(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): WARREN, Bridget, A. [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). BATRA, Sajeev [US/US]; 555 El Camino Real, #709, San Leandro, CA 94087 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). WALIA, Narinder, K. [US/US]; 890 David Street, #205, San Leandro, CA 94577 (US). ARVIZU, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). TANG, Y., Tom [US/US]; 4230 Ranwic Court, San Jose, CA 95118 (US). LU, Dyung, Aina, M. [US/US]; 233 Goy Drive, San Jose, CA 95123 (US). DUGGAN, Brendan, M. [AU/US]; 243 Buena Vista Avenue #306, Sunnyvale, CA 94086 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandero, CA 94577 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). KHAN, Farrah, A. [IN/US]; 9445 Harrison Street, Des Plaines, IL 60016 (US). NGUYEN, Danniel, B. [US/US]; 1403 Ridgwood Drive, San Jose, CA 95118 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). LAL, Preeti, G. [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). TRAN, Bao [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(54) Title: TRANSMEMBRANE PROTEINS

(57) Abstract: The invention provides human transmembrane proteins (TMP) and polynucleotides which identify and encode TMP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TMP.

TRANSMEMBRANE PROTEINS

TECHNICAL FIELD

5   This invention relates to nucleic acid and amino acid sequences of transmembrane proteins and to the use of these sequences in the diagnosis, treatment, and prevention of reproductive, developmental, cardiovascular, neurological, gastrointestinal, lipid metabolism, cell proliferative, and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transmembrane proteins.

10

BACKGROUND OF THE INVENTION

Eukaryotic organisms are distinct from prokaryotes in possessing many intracellular membrane-bound compartments such as organelles and vesicles. Many of the metabolic reactions which distinguish eukaryotic biochemistry from prokaryotic biochemistry take place within these

15   compartments. In particular, many cellular functions require very stringent reaction conditions, and the organelles and vesicles enable compartmentalization and isolation of reactions which might otherwise disrupt cytosolic metabolic processes. The organelles include mitochondria, smooth and rough endoplasmic reticula, sarcoplasmic reticulum, and the Golgi body. The vesicles include phagosomes, lysosomes, endosomes, peroxisomes, and secretory vesicles. Organelles and vesicles are bounded by

20   single or double membranes.

Biological membranes surround organelles, vesicles, and the cell itself. Membranes are highly selective permeability barriers made up of lipid bilayer sheets composed of phosphoglycerides, fatty acids, cholesterol, phospholipids, glycolipids, proteoglycans, and proteins. Membranes contain ion pumps, ion channels, and specific receptors for external stimuli which transmit biochemical signals

25   across the membranes. These membranes also contain second messenger proteins which interact with these pumps, channels, and receptors to amplify and regulate transmission of these signals.

Plasma Membrane Proteins

Transmembrane proteins (TMP) are characterized by extracellular, transmembrane, and intracellular domains. TMP domains are typically comprised of 15 to 25 hydrophobic amino acids

30   which are predicted to adopt an α-helical conformation. TMP are classified as bitopic (Types I and II) proteins, which span the membrane once, and polytopic (Types III and IV) (Singer, S.J. (1990) Annu. Rev. Cell Biol. 6:247-96) proteins, which contain multiple membrane-spanning segments. TMP that act as cell-surface receptor proteins involved in signal transduction include growth and differentiation factor receptors, and receptor-interacting proteins such as Drosophila pecanex and

35   frizzled proteins, LIV-1 protein, NF2 protein, and GNS1/SUR4 eukaryotic integral membrane

proteins. TMP also act as transporters of ions or metabolites, such as gap junction channels (connexins) and ion channels, and as cell anchoring proteins, such as lectins, integrins, and fibronectins. TMP function as vesicle and organelle-forming molecules, such as caveolins; or cell recognition molecules, such as cluster of differentiation (CD) antigens, glycoproteins, and mucins.

5   The transport of hydrophilic molecules across membranes is facilitated by the presence of channel proteins which form aqueous pores which can perforate a lipid bilayer. Many channels consist of protein complexes formed by the assembly of multiple subunits, at least one of which is an integral membrane protein that contributes to formation of the pore. In some cases, the pore is constructed to allow selective passage of only one or a few molecular species. Distinct types of

10   membrane channels that differ greatly in their distribution and selectivity include: (1) aquaporins, which transport water; (2) protein-conducting channels, which transport proteins across the endoplasmic reticulum membrane; (3) gap junctions, which facilitate diffusion of ions and small organic molecules between neighboring cells; and (4) ion channels, which regulate ion flux through various membranes.

15   Gap junctions (also called connexons) are specialized regions of the plasma membrane comprising transmembrane channels that function chemically and electrically to couple the cytoplasms of neighboring cells in many tissues. Gap junctions function as electrical synapses for intercellular propagation of action potentials in excitable tissues. In nonexcitable tissues, gap junctions have roles in tissue homeostasis, coordinated physiological response, metabolic cooperation, growth control, and the

20   regulation of development and differentiation.

Each connexon, which spans the lipid bilayer of the plasma membrane, is composed of six identical subunits called connexins. At least fourteen distinct connexin proteins exist, with each having similar structures but differing tissue distributions. Structurally, the connexins consist of a short cytoplasmic N-terminal domain connected to four transmembrane spanning regions (M1, M2, M3 and

25   M4) which separate two extracellular and one cytoplasmic loop followed by a C-terminal, cytoplasmic domain of variable length (20 resides in Cx26 to 260 residues in Cx56). The M2-M3 loop and the N- and C-termini are oriented towards the cell cytoplasm. Conserved regions include the membrane spanning regions and the two extracellular loops. Within the extracellular loops are three conserved cysteines which are involved in disulfide bond formation. Signature patterns for these two loops are

30   either: C-[DN]-T-x-Q-P-G-C-x-(2)-V-C-Y-D or C-x(3,4)-P-C-x(3)-[LIVM]-[DEN]-C-[FY]-[LIVM]-[SA]-[KR]-P (PDOC00341, Profilescan and S. Rahman and W.H. Evans, (1991) J. Cell Sci. 100:567-578). The variable regions, which include the cytoplasmic loop and the C-terminal region, may be responsible for the regulation of different connexins. (See Hennemann, H. et al. (1992) J. Biol. Chem. 267:17225-17233; PRINTS PR00206 connexin signature; Yeager, M. et al., (1998) Curr. O. Structr.

35   Biol. 8:517-524.)

Gap junctions help to synchronize heart and smooth muscle contraction, speed neural
transmission, and propagate extracellular signals. Gap junctions can open and close in response to
particular stimuli (e.g., pH, Ca^{+2}, and cAMP). The effective pore size of a gap junction is
approximately 1.5 nm, which enables small molecules (e.g., those under 1000 daltons) to diffuse freely

5   through the pore. Transported molecules include ions, small metabolites, and second messengers (e.g.,
Ca^{+2} and cAMP).

Connexins have many disease associations. Female mice lacking connexin 37 (Cx37) are
infertile due to the absence of the oocyte-granulosa cell signaling pathway. Mice lacking Cx43 die
shortly after birth and show cardiac defects reminiscent of some forms of stenosis of the pulmonary

10   artery in humans. Mutations in Cx32 are associated with the X-linked form of Charcot-Marie-Tooth
disease, a motor and sensory neuropathy of the peripheral nervous system. Cx26 is expressed in the
placenta, and Cx26-deficient mice show decreased transplacental transport of a glucose analog from the
maternal to the fetal circulation. In humans, Cx26 has been identified as the first susceptibility gene for
non-syndromic sensorineural autosomal deafness. Mutations in in Cx31 have been linked with an

15   autosomal-dominant hearing impairment (a nonsense or missense mutation in the second extracellular
loop) and in a dominantly transmitted skin disorder, erythrokeratodermia variabilis (missense mutations
in either the N-terminal domain or the M2 domain.) (See A. M. Simon, (1999) Trends Cell Biol. 9:169-
170). Cx46 is expressed in lens fiber cells, and Cx46-deficient mice develop early-onset cataracts that
resemble human nuclear cataracts. (See Nicholson, S.M. and R. Bruzzone (1997) Curr. Biol. 7:R340-

20   R344.)

Plasma membrane proteins (MPs) are divided into two groups based upon methods of protein
extraction from the membrane. Extrinsic or peripheral membrane proteins can be released using
extremes of ionic strength or pH, urea, or other disruptors of protein interactions. Intrinsic or integral
membrane proteins are released only when the lipid bilayer of the membrane is dissolved by

25   detergent.

Many membrane proteins (MPs) contain amino acid sequence motifs that serve to localize
proteins to specific subcellular sites. Examples of these motifs include PDZ domains, KDEL, RGD,
NGR, and GSL sequence motifs, von Willebrand factor A (vWFA) domains, and EGF-like domains.
RGD, NGR, and GSL motif-containing peptides have been used as drug delivery agents in targeted

30   cancer treatment of tumor vasculature (Arap, W. et al. (1998) Science, 279:377-380). Membrane
proteins may also contain amino acid sequence motifs that serve to interact with extracellular or
intracellular molecules, such as carbohydrate recognition domains.

Chemical modification of amino acid residue side chains alters the manner in which MPs
interact with other molecules, such as membrane phospholipids. Examples of such chemical

35   modifications include the formation of covalent bonds with glycosaminoglycans, oligosaccharides,

phospholipids, acetyl and palmitoyl moieties, ADP-ribose, phosphate, and sulphate groups.

RNA encoding membrane proteins may have alternative splice sites which give rise to proteins encoded by the same gene but with different messenger RNA and amino acid sequences. Splice variant membrane proteins may interact with other ligand and protein isoforms.

5   Transmembrane proteins of the plasma membrane also include cell surface receptors. These receptors recognize hormones such as catecholamines, e.g., epinephrine, norepinephrine, and histamine; peptide hormones, e.g., glucagon, insulin, gastrin, secretin, cholecystokinin, adrenocorticotropic hormone, follicle stimulating hormone, luteinizing hormone, thyroid stimulating hormone, parathyroid hormone, and vasopressin; growth and differentiation factors, e.g., epidermal

10   growth factor, fibroblast growth factor, transforming growth factor, insulin-like growth factor, platelet-derived growth factor, nerve growth factor, colony-stimulating factors, and erythropoietin; cytokines, e.g., chemokines, interleukins, interferons, and tumor necrosis factor; small peptide factors such as bombesin, oxytocin, endothelin, angiotensin II, vasoactive intestinal peptide, and bradykinin; neurotransmitters such as neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, e.g.,

15   enkephalins, endorphins and dynorphins; galanin, somatostatin, and tachykinins; and circulatory system-borne signaling molecules, e.g., angiotensin, complement, calcitonin, endothelins, and formyl-methionyl peptides. Cell surface receptors on immune system cells recognize antigens, antibodies, and major histocompatibility complex (MHC)-bound peptide. Other cell surface receptors bind ligands to be internalized by the cell. This receptor-mediated endocytosis functions in the uptake

20   of low density lipoproteins (LDL), transferrin, glucose- or mannose-terminal glycoproteins, galactose-terminal glycoproteins, immunoglobulins, phosphovitellogenins, fibrin, proteinase-inhibitor complexes, plasminogen activators, and thrombospondin. (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, p. 723; and Mikhailenko, I. et al. (1997) J. Biol. Chem. 272:6784-6791.)

25   Many cell surface receptors have seven transmembrane regions, with an extracellular N-terminus that binds ligand and a cytoplasmic C-terminus that interacts with G proteins. (Strosberg, A.D. (1991) Eur. J. Biochem. 196:1-10.) Cysteine-rich domains are found in two families of cell surface receptors, the LDL receptor family and the tumor necrosis factor receptor/nerve growth factor (TNFR/NGFR) receptor family. Seven successive cysteine-rich repeats of about forty amino acids in

30   the N-terminal extracellular region of the LDL receptor form the binding site for LDL and calcium; similar repeats have been found in vertebrate very low density lipoprotein receptor, vertebrate low-density lipoprotein receptor-related protein 1 (LRP1) (also known as α_2-macroglobulin receptor), and vertebrate low-density lipoprotein receptor-related protein 2 (also known as gp330 or megalin) (ExPASy PROSITE document PDOC00929; and Bairoch, A. et al. (1997) Nucl. Acids. Res. 25:217-

35   221.) The structure of the repeat is a β-hairpin followed by a series of β-turns; there are six disulfide-

bonded cysteines within each repeat.

The LDL receptor is an integral membrane protein which functions in lipid uptake by removing cholesterol from the blood. Most cells outside the liver and intestine take up cholesterol from the blood rather than synthesize it themselves. Cell surface LDL receptors bind LDL particles which are then

5   internalized by endocytosis (Meyers, R.A. (1995) Molecular Biology and Biotechnology, VCH Publishers, New York NY, pp. 494-501). Absence of the LDL receptor, the cause of the disease familial hypercholesterolemia, leads to increased plasma cholesterol levels and ultimately to atherosclerosis (Stryer, L. (1995) Biochemistry, W.H. Freeman, New York NY, pp. 691-702).

G-Protein Coupled Receptors

10   G-protein coupled receptors (GPCR) comprise a superfamily of integral membrane proteins which transduce extracellular signals. GPCRs include receptors for biogenic amines, lipid mediators of inflammation, peptide hormones, and sensory signal mediators.

The structure of these highly-conserved receptors consists of seven hydrophobic transmembrane (serpentine) regions, cysteine disulfide bridges between the second and third

15   extracellular loops, an extracellular N-terminus, and a cytoplasmic C-terminus. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. Cysteine disulfide bridges connect the second and third extracellular loops. A conserved, acidic-Arg-aromatic residue triplet present in the second cytoplasmic loop may interact with G

20   proteins. A GPCR consensus pattern is characteristic of most proteins belonging to this superfamily (ExPASy PROSITE document PS00237; and Watson, S. and S. Arkinstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego, CA, pp 2-6). Mutations and changes in transcriptional activation of GPCR-encoding genes have been associated with neurological disorders such as schizophrenia, Parkinson's disease, Alzheimer's disease, drug addiction, and feeding

25   disorders. The juvenile development and fertility-2 (jdf-2) locus, also called runty-jerky-sterile (rjs), is associated with deletions and point mutations in HERC2, a gene encoding a guanine nucleotide exchange factor protein involved in vesicular trafficking (Walkowicz, M. et al. (1999) Mamm. Genome 10:870-878).

A GPCR known as FP is the receptor for prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$). The prostaglandins

30   belong to a large family of naturally occurring paracrine/autocrine mediators of physiologic and inflammatory responses. $PGF_{2\alpha}$ plays a role in responses of certain tissues such as reproductive tract, lung, bone, and heart, including the stimulation of myometrial contraction, corpus luteum breakdown, and bronchoconstriction. An FP-associated molecule (FPRP) is copurified with FP and is expressed only in those tissues where a physiological role for $PGF_{2\alpha}$ has been described. FPRP is predicted to

35   be a transmembrane protein with glycosolated extracellular immunoglobulin loops and a short, highly

charged intracellular domain. FPRP appears to be a negative regulator of PGF$_{2\alpha}$ binding to FP. As such, FPRP may be associated with PGF$_{2\alpha}$ related diseases, which may include dysmenorrhea, infertility, asthma, or cardiomyophathy (Orlicky, D. J. et al. (1996) Hum. Genet. 97:655-658).

Scavenger Receptors

5   Macrophage scavenger receptors with broad ligand specificity may participate in the binding of low density lipoproteins (LDL) and foreign antigens. Scavenger receptors types I and II are trimeric membrane proteins with each subunit containing a small N-terminal intracellular domain, a transmembrane domain, a large extracellular domain, and a C-terminal cysteine-rich domain. The extracellular domain contains a short spacer domain, an α-helical coiled-coil domain, and a triple

10   helical collagenous domain. These receptors have been shown to bind a spectrum of ligands, including chemically modified lipoproteins and albumin, polyribonucleotides, polysaccharides, phospholipids, and asbestos (Matsumoto, A. et al. (1990) Proc. Natl. Acad. Sci. 87:9133-9137; and Elomaa, O. et al. (1995) Cell 80:603-609). The scavenger receptors are thought to play a key role in atherogenesis by mediating uptake of modified LDL in arterial walls, and in host defense by binding

15   bacterial endotoxins, bacteria, and protozoa.

Tetraspan family proteins

The transmembrane 4 superfamily (TM4SF), or tetraspan family, is a multigene family encoding type III integral membrane proteins (Wright, M.D. and Tomlinson, M.G. (1994) Immunol. Today 15:588-594). TM4SF is comprised of membrane proteins which traverse the cell membrane

20   four times. Members of the TM4SF include platelet and endothelial cell membrane proteins, melanoma-associated antigens, leukocyte surface glycoproteins, colonal carcinoma antigens, tumor-associated antigens, and surface proteins of the schistosome parasites (Jankowski, S.A. (1994) Oncogene 9:1205-1211). Members of the TM4SF share about 25-30% amino acid sequence identity with one another.

25   A number of TM4SF members have been implicated in signal transduction, control of cell adhesion, regulation of cell growth and proliferation, including development and oncogenesis, and cell motility, including tumor cell metastasis. Expression of TM4SF proteins is associated with a variety of tumors, and the level of expression may be altered when cells are growing or activated.

Tumor Antigens

30   Tumor antigens are surface molecules that are differentially expressed in tumor cells relative to normal cells. Tumor antigens distinguish tumor cells immunologically from normal cells and provide diagnostic and therapeutic targets for human cancers (Takagi, S. et al. (1995) Int. J. Cancer 61: 706-715; Liu, E. et al. (1992) Oncogene 7: 1027-1032).

Ion channels

35   Ion channels are found in the plasma membranes of virtually every cell in the body. For

example, chloride channels mediate a variety of cellular functions including regulation of membrane

potential and absorption and secretion of ions across epithelial membranes. When present in

intracellular membranes of the Golgi apparatus and endocytic vesicles, chloride channels also regulate

organelle pH (see, e.g., Greger, R. (1988) Annu. Rev. Physiol. 50:111-122). Electrophysiological and

5 pharmacological properties of chloride channels, including ion conductance, current-voltage

relationships, and sensitivity to modulators, suggest that different chloride channels exist in muscles,

neurons, fibroblasts, epithelial cells, and lymphocytes.

Many channels have sites for phosphorylation by one or more protein kinases including

protein kinase A, protein kinase C, casein kinase II, and tyrosine kinases, all of which regulate ion

10 channel activity in cells. Inappropriate phosphorylation of membrane proteins has been correlated

with pathological changes in cell cycle progression and cell differentiation. Changes in the cell cycle

have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked

to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

Cerebellar granule neurons possess a non-inactivating potassium current which modulates

15 firing frequency upon receptor stimulation by neurotransmitters and controls the resting membrane

potential. Potassium channels that exhibit non-inactivating currents include the *ether a go-go* (EAG)

channel. A membrane protein designated KCR1 specifically binds to rat EAG by means of its C-

terminal region and regulates the cerebellar non-inactivating potassium current. KCR1 is predicted to

contain 12 transmembrane domains, with intracellular amino and carboxyl termini. Structural

20 characteristics of these transmembrane regions appear to be similar to those of the transporter

superfamily, but no homology between KCR1 and known transporters was found, suggesting that

KCR1 belongs to a novel class of transporters. KCR1 appears to be the regulatory component of non-

inactivating potassium channels (Hoshi, N. et al. (1998) J. Biol. Chem. 273:23080-23085).

Proton pumps

25 Proton ATPases are a large class of membrane proteins that use the energy of ATP hydrolysis

to generate an electrochemical proton gradient across a membrane. The resultant gradient may be

used to transport other ions across the membrane (Na^+, K^+, or Cl^-) or to maintain organelle pH.

Proton ATPases are further subdivided into the mitochondrial F-ATPases, the plasma membrane

ATPases, and the vacuolar ATPases. The vacuolar ATPases establish and maintain an acidic pH

30 within various vesicles involved in the processes of endocytosis and exocytosis (Mellman, I. et al.

(1986) Ann. Rev. Biochem. 55:663-700).

Proton-coupled, 12 membrane-spanning domain transporters such as PEPT 1 and PEPT 2 are

responsible for gastrointestinal absorption and for renal reabsorption of peptides using an

electrochemical H^+ gradient as the driving force. Another type of peptide transporter, the TAP

35 transporter, is a heterodimer consisting of TAP 1 and TAP 2 and is associated with antigen

processing. Peptide antigens are transported across the membrane of the endoplasmic reticulum by TAP so they can be expressed on the cell surface in association with MHC molecules. Each TAP protein consists of multiple hydrophobic membrane spanning segments and a highly conserved ATP-binding cassette (Boll, M. et al. (1996) Proc. Natl. Acad. Sci. 93:284-289). Pathogenic

5   microorganisms, such as herpes simplex virus, may encode inhibitors of TAP-mediated peptide transport in order to evade immune surveillance (Marusina, K. and Manaco, J.J. (1996) Curr. Opin. Hematol. 3:19-26).

ABC Transporters

The ATP-binding cassette (ABC) transporters, also called the "traffic ATPases," comprise a

10   superfamily of membrane proteins that mediate transport and channel functions in prokaryotes and eukaryotes (Higgins, C.F. (1992) Annu. Rev. Cell Biol. 8:67-113). ABC proteins share a similar overall structure and significant sequence homology. All ABC proteins contain a conserved domain of approximately two hundred amino acid residues which includes one or more nucleotide binding domains. Mutations in ABC transporter genes are associated with various disorders, such as

15   hyperbilirubinemia II/Dubin-Johnson syndrome, recessive Stargardt's disease, X-linked adrenoluekodystrophy, multidrug resistance, celiac disease, and cystic fibrosis.

Membrane Proteins Associated with Intercellular Communication

Intercellular communication is essential for the development and survival of multicellular organisms. Cells communicate with one another through the secretion and uptake of protein signaling

20   molecules. The uptake of proteins into the cell is achieved by endocytosis, in which the interaction of signaling molecules with the plasma membrane surface, often via binding to specific receptors, results in the formation of plasma membrane-derived vesicles that enclose and transport the molecules into the cytosol. The secretion of proteins from the cell is achieved by exocytosis, in which molecules inside of the cell are packaged into membrane-bound transport vesicles derived from the *trans* Golgi network.

25   These vesicles fuse with the plasma membrane and release their contents into the surrounding extracellular space. Endocytosis and exocytosis result in the removal and addition of plasma membrane components, and the recycling of these components is essential to maintain the integrity, identity, and functionality of both the plasma membrane and internal membrane-bound compartments.

Synaptobrevins are synaptic vesicle-associated membrane proteins (VAMPs) which were first

30   discovered in rat brain. These proteins were initially thought to be limited to neuronal cells and to function in the movement of vesicles from the plasmalemma of one cell, across the synapse, to the plasmalemma of another cell. Synaptobrevins are now known to occur and function in constitutive vesicle trafficking pathways involving receptor-mediated endocytotic and exocytotic pathways of many non-neuronal cell types. This regulated vesicle trafficking pathway may be blocked by the highly

35   specific action of clostridial neurotoxins which cleave the synaptobrevin molecule.

In vitro studies of various cellular membranes (Galli et al. (1994) J. Cell. Biol. 125:1015-24; Link et al. (1993) J. Biol. Chem. 268:18423-6) have shown that VAMPS are widely distributed. These important membrane trafficking proteins appear to participate in axon extension via exocytosis during development, in the release of neurotransmitters and modulatory peptides, and in endocytosis.

5   Endocytotic vesicular transport includes such intracellular events as the fusions and fissions of the nuclear membrane, endoplasmic reticulum, Golgi apparatus, and various inclusion bodies such as peroxisomes or lysosomes. Endocytotic processes appear to be universal in eukaryotic cells as diverse as yeast, Caenorhabditis elegans, Drosophila, and mammals.

VAMP-1B is involved in subcellular targeting and is an isoform of VAMP-1A (Isenmann, S. et

10   al. (1998) Mol. Biol. Cell 9:1649-1660). Four additional splice variants (VAMP-1C to F) have recently been identified. Each variant has variable sequences only at the extreme C-terminus, suggesting that the C-terminus is important in vesicle targeting (Berglund, L. et al. (1999) Biochem. Biophys. Res. Commun. 264:777-780).

Lysosomes are the site of degradation of intracellular material during autophagy, and of

15   extracellular molecules following endocytosis. Lysosomal enzymes are packaged into vesicles which bud from the trans-Golgi network. These vesicles fuse with endosomes to form the mature lysosome in which hydrolytic digestion of endocytosed material occurs. Lysosomes can fuse with autophagosomes to form a unique compartment in which the degradation of organelles and other intracellular components occurs.

20   Protein sorting by transport vesicles, such as the endosome, has important consequences for a variety of physiological processes including cell surface growth, the biogenesis of distinct intracellular organelles, endocytosis, and the controlled secretion of hormones and neurotransmitters (Rothman, J.E. and Wieland, F.T. (1996) Science 272:227-234). In particular, neurodegenerative disorders and other neuronal pathologies are associated with biochemical flaws during endosomal protein sorting or

25   endosomal biogenesis (Mayer R.J. et al. (1996) Adv. Exp. Med. Biol. 389:261-269).

Peroxisomes are organelles independent from the secretory pathway. They are the site of many peroxide-generating oxidative reactions in the cell. Peroxisomes are unique among eukaryotic organelles in that their size, number, and enzyme content vary depending upon organism, cell type, and metabolic needs (Waterham, H.R. and Cregg, J.M. (1997) BioEssays 19:57-66). Genetic defects in

30   peroxisome proteins which result in peroxisomal deficiencies have been linked to a number of human pathologies, including Zellweger syndrome, rhizomelic chondrodysplasia punctata, X-linked adrenoleukodystrophy, acyl-CoA oxidase deficiency, bifunctional enzyme deficiency, classical Refsum's disease, DHAP alkyl transferase deficiency, and acatalasemia (Moser, H.W. and Moser, A.B. (1996) Ann. NY Acad. Sci. 804:427-441). In addition, Gartner, J. et al. (1991; Pediatr. Res. 29:141-

35   146) found a 22 kDa integral membrane protein associated with lower density peroxisome-like

subcellular fractions in patients with Zellweger syndrome.

Normal embryonic development and control of germ cell maturation is modulated by a number of secretory proteins which interact with their respective membrane-bound receptors. Cell fate during embryonic development is determined by members of the activin/TGF-β superfamily, cadherins, IGF-
5   2, and other morphogens. In addition, proliferation, maturation, and redifferentiation of germ cell and reproductive tissues are regulated, for example, by IGF-2, inhibins, activins, and follistatins (Petraglia, F. (1997) Placenta 18:3-8; Mather, J.P. et al. (1997) Proc. Soc. Exp. Biol. Med. 215:209-222).

Endoplasmic Reticulum Membrane Proteins

10   The normal functioning of the eukaryotic cell requires that all newly synthesized proteins be correctly folded, modified, and delivered to specific intra- and extracellular sites. Newly synthesized membrane and secretory proteins enter a cellular sorting and distribution network during or immediately after synthesis and are routed to specific locations inside and outside of the cell. The initial compartment in this process is the endoplasmic reticulum (ER) where proteins undergo
15   modifications such as glycosylation, disulfide bond formation, and oligomerization. The modified proteins are then transported through a series of membrane-bound compartments which include the various cisternae of the Golgi complex, where further carbohydrate modifications occur. Transport between compartments occurs by means of vesicle budding and fusion. Once within the secretory pathway, proteins do not have to cross a membrane to reach the cell surface.

20   Although the majority of proteins processed through the ER are transported out of the organelle, some are retained. The signal for retention in the ER in mammalian cells consists of the tetrapeptide sequence, KDEL, located at the carboxyl terminus of resident ER membrane proteins (Munro, S. (1986) Cell 46:291-300). Proteins containing this sequence leave the ER but are quickly retrieved from the early Golgi cisternae and returned to the ER, while proteins lacking this signal
25   continue through the secretory pathway.

Disruptions in the cellular secretory pathway have been implicated in several human diseases. In familial hypercholesterolemia the low density lipoprotein receptors remain in the ER, rather than moving to the cell surface (Pathak, R.K. (1988) J. Cell Biol. 106:1831-1841). Altered transport and processing of the β-amyloid precursor protein (βAPP) involves the putative vesicle transport protein
30   presenilin and may play a role in early-onset Alzheimer's disease (Levy-Lahad, E. et al. (1995) Science 269:973-977). Changes in ER-derived calcium homeostasis have been associated with diseases such as cardiomyopathy, cardiac hypertrophy, myotonic dystrophy, Brody disease, Smith-McCort dysplasia, and diabetes mellitus.

Mitochondrial Membrane Proteins

35   The mitochondrial electron transport (or respiratory) chain is a series of three enzyme

complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP then provides the primary source of energy for driving the many energy-requiring reactions of a cell.

5   Most of the protein components of the mitochondrial respiratory chain are the products of nuclear encoded genes that are imported into the mitochondria, and the remainder are products of mitochondrial genes. Defects and altered expression of enzymes in the respiratory chain are associated with a variety of disease conditions in man, including, for example, neurodegenerative diseases, myopathies, and cancer.

10   **Lymphocyte and Leukocyte Membrane Proteins**

The B-cell response to antigens is an essential component of the normal immune system. Mature B cells recognize foreign antigens through B cell receptors (BCR) which are membrane-bound, specific antibodies that bind foreign antigens. The antigen/receptor complex is internalized, and the antigen is proteolytically processed. To generate an efficient response to complex antigens,

15   the BCR, BCR-associated proteins, and T cell response are all required. Proteolytic fragments of the antigen are complexed with major histocompatability complex-II (MHCII) molecules on the surface of the B cells where the complex can be recognized by T cells. In contrast, macrophages and other lymphoid cells present antigens in association with MHCI molecules to T cells. T cells recognize and are activated by the MHCI-antigen complex through interactions with the T cell receptor/CD3

20   complex, a T cell-surface multimeric protein located in the plasma membrane. T cells activated by antigen presentation secrete a variety of lymphokines that induce B cell maturation and T cell proliferation, and activate macrophages, which kill target cells.

Leukocytes have a fundamental role in the inflammatory and immune response, and include monocytes/macrophages, mast cells, polymorphonucleoleukocytes, natural killer cells, neutrophils,

25   eosinophils, basophils, and myeloid precursors. Leukocyte membrane proteins include members of the CD antigens, N-CAM, I-CAM, human leukocyte antigen (HLA) class I and HLA class II gene products, immunoglobulins, immunoglobulin receptors, complement, complement receptors, interferons, interferon receptors, interleukin receptors, and chemokine receptors.

Abnormal lymphocyte and leukocyte activity has been associated with acute disorders such as

30   AIDS, immune hypersensitivity, leukemias, leukopenia, systemic lupus, granulomatous disease, and eosinophilia.

Apoptosis-Associated Membrane Proteins

A variety of ligands, receptors, enzymes, tumor suppressors, viral gene products, pharmacological agents, and inorganic ions have important positive or negative roles in regulating

35   and implementing the apoptotic destruction of a cell. Although some specific components of the

apoptotic pathway have been identified and characterized, many interactions between the proteins involved are undefined, leaving major aspects of the pathway unknown.

A requirement for calcium in apoptosis was previously suggested by studies showing the involvement of calcium levels in DNA cleavage and Fas-mediated cell death (Hewish, D.R. and L.A.

5   Burgoyne (1973) Biochem. Biophys. Res. Comm. 52:504-510; Vignaux, F. et al. (1995) J. Exp. Med. 181:781-786; Oshimi, Y. and S. Miyazaki (1995) J. Immunol. 154:599-609). Other studies show that intracellular calcium concentrations increase when apoptosis is triggered in thymocytes by either T cell receptor cross-linking or by glucocorticoids, and cell death can be prevented by blocking this increase (McConkey, D.J. et al. (1989) J. Immunol. 143:1801-1806; McConkey, D.J. et al. (1989)

10   Arch. Biochem. Biophys. 269:365-370). Therefore, membrane proteins such as calcium channels and the Fas receptor are important for the apoptotic response.

The discovery of new transmembrane proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of reproductive, developmental, cardiovascular, neurological, gastrointestinal, lipid

15   metabolism, cell proliferative, and autoimmune/inflammatory disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transmembrane proteins.

SUMMARY OF THE INVENTION

20   The invention features purified polypeptides, transmembrane proteins, referred to collectively as "TMP" and individually as "TMP-1," "TMP-2," "TMP-3," "TMP-4," "TMP-5," "TMP-6," "TMP-7," "TMP-8," "TMP-9," "TMP-10," "TMP-11," "TMP-12," "TMP-13," "TMP-14," "TMP-15," "TMP-16," and "TMP-17." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group

25   consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the

30   invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-17.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-

35   17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the

12

group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-17. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:18-34.

5   Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically
10   active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

15   The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group
20   consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

25   Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid
30   sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90%
35   identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a

polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

5   Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the

10   polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of

15   said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID

20   NO:18-34, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b)

25   detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid

30   sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid

35   sequence selected from the group consisting of SEQ ID NO:1-17. The invention additionally provides a

method of treating a disease or condition associated with decreased expression of functional TMP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

5   acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID

10   NO:1-17. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TMP, comprising administering to a

15   patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence

20   selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the

25   invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TMP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds

30   to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic

35   fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID

15

NO:1-17. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

    The invention further provides a method of screening for a compound that modulates the

5   activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17, and d) an immunogenic

10 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-17. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound,

15 wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

    The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, the method

20 comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

    The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20

25 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the

30 polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90%

35 identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:18-34, iii) a

polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization · complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a

17

reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although

5   any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is

10   not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TMP" refers to the amino acid sequences of substantially purified TMP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

15   The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TMP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TMP either by directly interacting with TMP or by acting on components of the biological pathway in which TMP participates.

An "allelic variant" is an alternative form of the gene encoding TMP. Allelic variants may

20   result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in

25   a given sequence.

"Altered" nucleic acid sequences encoding TMP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TMP or a polypeptide with at least one functional characteristic of TMP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of

30   the polynucleotide encoding TMP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TMP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TMP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,

35   hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological

or immunological activity of TMP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains

5   having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring

10   protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

15   The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TMP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TMP either by directly interacting with TMP or by acting on components of the biological pathway in which TMP participates.

20   The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')$_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TMP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or

25   synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to

30   immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific

35   molecular target. Aptamers are derived from an _in vitro_ evolutionary process (e.g., SELEX

(Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.

5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries.

Aptamer compositions may be double-stranded or single-stranded, and may include

deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The

5 nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a

ribonucleotide may be replaced by 2'-F or 2'-NH$_2$), which may improve a desired property, e.g.,

resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules,

e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system.

Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-

10 linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed _in vivo_. For example, a vaccinia

virus-based RNA expression system has been used to express specific RNA aptamers at high levels in

the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-

15 handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed

nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on

substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense"

(coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA;

20 peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as

phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified

sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having

modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense

molecules may be produced by any method including chemical synthesis or transcription. Once

25 introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring

nucleic acid sequence produced by the cell to form duplexes which block either transcription or

translation. The designation "negative" or "minus" can refer to the antisense strand, and the

designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical

30 functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"

refers to the capability of the natural, recombinant, or synthetic TMP, or of any oligopeptide thereof, to

induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid

sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,

35 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TMP or fragments of TMP may be

5   employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated

10   DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to

15   produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as

20   conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

5 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one

10 biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

15 "Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon

20 may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of TMP or the polynucleotide encoding TMP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to

25 the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For

30 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

35 A fragment of SEQ ID NO:18-34 comprises a region of unique polynucleotide sequence that

22

specifically identifies SEQ ID NO:18-34, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:18-34 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:18-34 from related polynucleotide sequences. The precise length of a fragment of SEQ ID

5   NO:18-34 and the region of SEQ ID NO:18-34 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-17 is encoded by a fragment of SEQ ID NO:18-34. A fragment of SEQ ID NO:1-17 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-17. For example, a fragment of SEQ ID NO:1-17 is useful as an immunogenic peptide for the

10   development of antibodies that specifically recognize SEQ ID NO:1-17. The precise length of a fragment of SEQ ID NO:1-17 and the region of SEQ ID NO:1-17 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon

15   (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to

20   the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default

25   parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows:

30   Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search

35   Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several

sources, including the NCBI, Bethesda, MD, and on the Internet at
http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis
programs including "blastn," that is used to align a known polynucleotide sequence with other
polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2
5   Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2
Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The
"BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST
programs are commonly used with gap and other parameters set to default settings. For example, to
compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version
10   2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

> *Matrix: BLOSUM62*
>
> *Reward for match: 1*
>
> *Penalty for mismatch: -2*
>
> *Open Gap: 5 and Extension Gap: 2 penalties*
>
> 15   *Gap x drop-off: 50*
>
> *Expect: 10*
>
> *Word Size: 11*
>
> *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as
20   defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over
the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at
least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such
lengths are exemplary only, and it is understood that any fragment length supported by the sequences
shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which
25   percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode
similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in
a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences
that all encode substantially the same protein.

30   The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to
the percentage of residue matches between at least two polypeptide sequences aligned using a
standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment
methods take into account conservative amino acid substitutions. Such conservative substitutions,
explained in more detail above, generally preserve the charge and hydrophobicity at the site of
35   substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and

5   "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12

10   (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

15   *Word Size: 3*

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance,

20   a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain

25   DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

30   "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e.,

35

25

binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive

5   annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic

10   strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

15   High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance,

20   sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is

25   strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g.,

30   paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

35   disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of

various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TMP which is capable of eliciting an immune response when introduced into a living organism, for example, a

5   mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TMP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other

10 chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TMP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TMP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide,

15 polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably

20 linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of

25 amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TMP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the

30 art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TMP.

"Probe" refers to nucleic acid sequences encoding TMP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels

35 include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short

27

nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

5   Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the
10   specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR
15   Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such
20   purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South
25   West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for
30   microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences.
35   Hence, this program is useful for identification of both unique and conserved oligonucleotides and

polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to
5   those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques
10   such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a
15   vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA
20   stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

25   An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TMP,
30   nucleic acids encoding TMP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or
35   synthetic binding composition. The interaction is dependent upon the presence of a particular structure

29

of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

5   The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by
10   different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

15   A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences
20   into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed
25   cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor
30   of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be
35   introduced into the host by methods known in the art, for example infection, transfection,

transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

5   A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence

10   identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule.

15   Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be

20   indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least

25   60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

30   The invention is based on the discovery of new human transmembrane proteins (TMP), the polynucleotides encoding TMP, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive, developmental, cardiovascular, neurological, gastrointestinal, lipid metabolism, cell proliferative, and autoimmune/inflammatory disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide

35   sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a

31

single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an

5   Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3

10   shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and

15   2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI).

20   Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transmembrane proteins. For example, SEQ ID

25   NO:2 is 89% identical to rat prostaglandin F2a receptor regulatory protein (GenBank ID g1054884) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains six immunoglubulin domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM

30   database of conserved protein family domains. (See Table 3.) In addition, SEQ ID NO:2 contains a signal peptide, a transmembrane domain, and an RGD motif, providing further corroborative evidence that SEQ ID NO:2 is a human transmembrane protein.

In the alternative, SEQ ID NO:4 is 56% identical to human connexin 31.1 (GenBank ID g4336903) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The

35   BLAST probability score is 5.8e-68, which indicates the probability of obtaining the observed

polypeptide sequence alignment by chance. SEQ ID NO:4 also contains a connexin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a connexin.

5   Note that six identical connexins compose a connexon (gap junction), a transmembrane channel in the plasma membrane which functions chemically and electrically to couple the cytoplasms of neighboring cells in many tissues. SEQ ID NO:5 is 1554 amino acids in length and is 99% identical over 1157 amino acids to human MEGF7 (GenBank ID g3449306) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the

10   probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains low-density lipoprotein receptor repeats and low-density lipoprotein receptor domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:5 is a member of the LDL receptor family of

15   proteins.

        In another alternative, SEQ ID NO:6 is 36% identical to mouse low density lipoprotein receptor related protein LRP1B/LRP-DIT (GenBank ID g8926243) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-40, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also

20   contains low-density lipoprotein receptor domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:6 is a low-density lipoprotein receptor-related molecule.   Further, SEQ ID NO:14 is 59% identical to human TNF-inducible protein CG12-1 (GenBank ID g3978246) as

25   determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.2e-94, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from HMMER analysis provides further corroborative evidence that SEQ ID NO:14 contains a transmembrane domain. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:7-13, and SEQ ID NO:15-17 were analyzed and annotated in a similar manner. The algorithms and

30   parameters for the analysis of SEQ ID NO:1-17 are described in Table 7.

        As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide

35   consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.

Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:18-34 or that distinguish between SEQ ID NO:18-34 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA

5   sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

10   The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6798827J1 is the identification number of an Incyte cDNA sequence, and COLENOR03 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71760758V1). Alternatively, the identification numbers in column 5 may refer to

15   GenBank cDNAs or ESTs (e.g., g1506355) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (i.e., those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (i.e., those sequences including

20   the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (i.e., those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_$XXXXXX\_N_1\_N_2\_YYYYY\_N_3\_N_4$ represents a "stitched" sequence in which $XXXXXX$ is the identification number of the cluster of sequences to which the algorithm was

25   applied, and $YYYYY$ is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL$XXXXXX\_gAAAAA\_gBBBBB\_1\_N$ is the identification number of a "stretched" sequence, with $XXXXXX$ being the Incyte project identification

30   number, g$AAAAA$ being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g$BBBBB$ being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be

35   used in place of the GenBank identifier (i.e., g$BBBBB$).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TMP variants. A preferred TMP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TMP amino acid sequence, and which contains at least one functional or structural characteristic of TMP.

The invention also encompasses polynucleotides which encode TMP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34, which encodes TMP. The polynucleotide sequences of SEQ ID NO:18-34, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TMP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least

about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TMP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:18-34 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide

5   sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:18-34. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TMP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding TMP. A splice variant may have portions which have significant

10   sequence identity to the polynucleotide sequence encoding TMP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding TMP over its entire length; however, portions of the

15   splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding TMP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TMP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic

20   code, a multitude of polynucleotide sequences encoding TMP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally

25   occurring TMP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TMP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TMP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TMP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring

30   codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TMP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from

35   the naturally occurring sequence.

36

The invention also encompasses production of DNA sequences which encode TMP and TMP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into
5   a sequence encoding TMP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:18-34 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol.
10   152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied
15   Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler
20   (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular
25   Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TMP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic
30   DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent
35   to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al.

37

(1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060).

5   Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a

10   GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library

15   does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-

20   specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments

25   which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TMP may be cloned in recombinant DNA molecules that direct expression of TMP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent

30   amino acid sequence may be produced and used to express TMP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TMP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic

35   oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-

mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen, Santa Clara CA; described in U.S. Patent No. 5,837,458;

5   Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TMP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then

10   subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are

15   optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TMP may be synthesized, in whole or in part,

20   using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TMP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.

25   55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TMP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

30   The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TMP, the nucleotide sequences encoding TMP or

35   derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains

the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TMP. Such elements may vary in their strength and specificity. Specific initiation signals

5   may also be used to achieve more efficient translation of sequences encoding TMP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TMP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous

10   translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

15   Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TMP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995)

20   Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TMP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus);

25   plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO

30   J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di

35   Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci.

USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

        In bacterial systems, a number of cloning and expression vectors may be selected depending

5   upon the use intended for polynucleotide sequences encoding TMP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TMP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TMP into the vector's multiple cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed

10   bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TMP are needed, e.g. for the production of antibodies, vectors which direct high level expression of TMP may be used. For example, vectors containing the

15   strong, inducible SP6 or T7 bacteriophage promoter may be used.

        Yeast expression systems may be used for production of TMP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into

20   the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

        Plant systems may also be used for expression of TMP. Transcription of sequences encoding TMP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). ·

25   Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp.

30   191-196.)

        In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TMP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain

35   infective virus which expresses TMP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc.

Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
5   DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TMP in cell lines is preferred. For example, sequences encoding TMP can be transformed into cell
10   lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the
15   introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232;
20   Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.)
25   Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of
30   transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TMP is inserted within a marker gene sequence, transformed cells containing
35   sequences encoding TMP can be identified by the absence of marker gene function. Alternatively, a

42

marker gene can be placed in tandem with a sequence encoding TMP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TMP and that express
5   TMP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TMP using either
10   specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TMP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al.
15   (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization
20   or PCR probes for detecting sequences related to polynucleotides encoding TMP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TMP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6
25   and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

30   Host cells transformed with nucleotide sequences encoding TMP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TMP may be designed to contain signal sequences which direct secretion
35   of TMP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the

5   protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

10   In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TMP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TMP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TMP activity. Heterologous protein and peptide moieties

15   may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,

20   respectively. FLAG, c-myc, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TMP encoding sequence and the heterologous protein sequence, so that TMP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein

25   expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TMP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or

30   SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

TMP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TMP. At least one and up to a plurality of test compounds may be screened for specific binding to TMP. Examples of test compounds include antibodies, oligonucleotides, proteins

35   (e.g., receptors), or small molecules.

44

In one embodiment, the compound thus identified is closely related to the natural ligand of TMP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TMP

5   binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TMP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TMP or cell membrane fractions which contain TMP are then contacted with a

10   test compound and binding, stimulation, or inhibition of activity of either TMP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TMP, either in

15   solution or affixed to a solid support, and detecting the binding of TMP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

20   TMP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TMP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TMP activity, wherein TMP is combined with at least one test compound, and the activity of TMP in the presence of a test compound is compared with the activity of TMP in the absence of the test compound.

25   A change in the activity of TMP in the presence of the test compound is indicative of a compound that modulates the activity of TMP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TMP under conditions suitable for TMP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TMP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test

30   compounds may be screened.

In another embodiment, polynucleotides encoding TMP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example,

35   mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and

45

grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP

5   system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.

10 Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TMP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998)

15 Science 282:1145-1147).

Polynucleotides encoding TMP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TMP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted

20 as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TMP, e.g., by secreting TMP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

25 Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TMP and transmembrane proteins. In addition, the expression of TMP is closely associated with brain, prostate, smooth muscle, cardiovascular, pituitary, gastrointestinal, lung, pancreatic, and small intestine tissues. Therefore, TMP appears to play a role in reproductive, developmental, cardiovascular, neurological, gastrointestinal, lipid metabolism, cell proliferative, and

30 autoimmune/inflammatory disorders. In the treatment of disorders associated with increased TMP expression or activity, it is desirable to decrease the expression or activity of TMP. In the treatment of disorders associated with decreased TMP expression or activity, it is desirable to increase the expression or activity of TMP.

Therefore, in one embodiment, TMP or a fragment or derivative thereof may be administered

35 to a subject to treat or prevent a disorder associated with decreased expression or activity of TMP.

46

Examples of such disorders include, but are not limited to, a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders,

5   ectopic pregnancy, teratogenesis, cancer of the breast, fibrocystic breast disease, galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and

10   anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome,

15   hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease,

20   aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse,

25   rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis,

30   obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia,

35   diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary

47

hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar

proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax,

pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung

transplantation; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke,

5   cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's

disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron

disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple

sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural

empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral

10   central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and

Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases

of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic

15   nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis,

inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental

disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses,

20   postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration,

and familial frontotemporal dementia; a gastrointestinal disorder such as dysphagia, peptic esophagitis,

esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric

carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis,

gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis,

25   cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis,

hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative

colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic

carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation,

gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice,

30   hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease,

alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal

vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis,

veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic

cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and

35   carcinomas; a lipid metabolism disorder such as fatty liver, cholestasis, primary biliary cirrhosis,

48

carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus,

5   lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease,

10   hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone,

15   bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia,

20   autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel

25   syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal,

30   parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing TMP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TMP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TMP in conjunction

35   with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder

associated with decreased expression or activity of TMP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TMP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity
5   of TMP including, but not limited to, those listed above.

In a further embodiment, an antagonist of TMP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TMP. Examples of such disorders include, but are not limited to, those reproductive, developmental, cardiovascular, neurological, gastrointestinal, lipid metabolism, cell proliferative, and autoimmune/inflammatory
10 disorders described above. In one aspect, an antibody which specifically binds TMP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TMP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TMP may be administered to a subject to treat or prevent a disorder associated with increased
15 expression or activity of TMP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination
20 of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TMP may be produced using methods which are generally known in the art. In particular, purified TMP may be used to produce antibodies or to screen libraries of pharmaceutical
25 agents to identify those which specifically bind TMP. Antibodies to TMP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

30   For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TMP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols,
35   polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG

50

(bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TMP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are

5   identical to a portion of the amino acid sequence of the natural protein. Short stretches of TMP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TMP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited

10 to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the

15 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TMP-specific single chain

20 antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in

25 the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TMP may also be generated. For example, such fragments include, but are not limited to, F(ab')$_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2

30 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either

35 polyclonal or monoclonal antibodies with established specificities are well known in the art. Such

immunoassays typically involve the measurement of complex formation between TMP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TMP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

5   Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TMP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of TMP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TMP

10   epitopes, represents the average affinity, or avidity, of the antibodies for TMP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TMP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TMP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7

15   L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TMP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

       The titer and avidity of polyclonal antibody preparations may be further evaluated to determine

20   the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TMP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and

25   Coligan et al. supra.)

       In another embodiment of the invention, the polynucleotides encoding TMP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TMP.

30   Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TMP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ.)

       In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered

35   intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence

complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood

5   76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

10   In another embodiment of the invention, polynucleotides encoding TMP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

15   (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii)

20   express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides

25   brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TMP expression or regulation causes disease, the expression of TMP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

        In a further embodiment of the invention, diseases or disorders caused by deficiencies in TMP

30   are treated by constructing mammalian expression vectors encoding TMP and introducing these vectors by mechanical means into TMP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics,

35   Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TMP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TMP may be

5   expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the

10   ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TMP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

15   TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these

20   standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TMP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TMP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive

25   element (RRE) along with additional retrovirus cis-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

30   receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a

35   method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4[+] T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al.

5   (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TMP to cells which have one or more genetic abnormalities with respect to the expression of TMP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile

10   for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

15   In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TMP to target cells which have one or more genetic abnormalities with respect to the expression of TMP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TMP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with

20   ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which

25   consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the

30   generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TMP to target cells. The biology of the prototypic alphavirus, Semliki

35   Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV

genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease

5 and polymerase). Similarly, inserting the coding sequence for TMP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TMP-coding RNAs and the synthesis of high levels of TMP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication

10 of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TMP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus

15 infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases,

20 transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

25 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TMP.

30 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of

35 candidate targets may also be evaluated by testing accessibility to hybridization with complementary

oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

5   Alternatively, RNA molecules may be generated by _in vitro_ and _in vivo_ transcription of DNA sequences encoding TMP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

10   modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine,

15   guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TMP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides,

20   transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TMP expression or activity, a compound which specifically inhibits expression of the polynucleotide

25   encoding TMP may be therapeutically useful, and in the treatment of disorders associated with decreased TMP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TMP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

30   commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a

35   polynucleotide encoding TMP is exposed to at least one test compound thus obtained. The sample

may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TMP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the

5   polynucleotide encoding TMP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide

10   can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and

15   modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

20   Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and

25   monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's

30   Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TMP, antibodies to TMP, and mimetics, agonists, antagonists, or inhibitors of TMP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,

35   sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins),

5   recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active

10   ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TMP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the

15   macromolecule. Alternatively, TMP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell

20   culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TMP or

25   fragments thereof, antibodies of TMP, and agonists, antagonists or inhibitors of TMP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the

30   therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the

35   patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time

5   and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and

10 methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

15 In another embodiment, antibodies which specifically bind TMP may be used for the diagnosis of disorders characterized by expression of TMP, or in assays to monitor patients being treated with TMP or agonists, antagonists, or inhibitors of TMP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TMP include methods which utilize the antibody and a label to detect TMP in human body fluids or in extracts of

20 cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TMP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TMP expression. Normal or

25 standard values for TMP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TMP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TMP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation

30 between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TMP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TMP may be correlated with

35 disease. The diagnostic assay may be used to determine absence, presence, and excess expression of

TMP, and to monitor regulation of TMP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TMP or closely related molecules may be used to identify nucleic acid sequences which encode TMP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TMP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TMP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:18-34 or from genomic sequences including promoters, enhancers, and introns of the TMP gene.

Means for producing specific hybridization probes for DNAs encoding TMP include the cloning of polynucleotide sequences encoding TMP or TMP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TMP may be used for the diagnosis of disorders associated with expression of TMP. Examples of such disorders include, but are not limited to, a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis, cancer of the breast, fibrocystic breast disease, galactorrhea, a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary

neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism,

hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida,

anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a

cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis,

5   Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and

phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular

replacement, coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease,

angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease,

calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral

10   valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial

thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease,

cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease,

complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary

congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction,

15   pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary

disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma,

bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary

tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis,

desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia

20   bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes,

Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in

collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and

noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease,

radiation-induced lung disease, and complications of lung transplantation; a neurological disorder

25   such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease,

Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal

disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural

muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating

diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess,

30   suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system

disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-

Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the

nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis,

encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central

35   nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic

nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

5 akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis,

10 gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation,

15 gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic

20 cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a lipid metabolism disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM$_2$ gangliosidosis, and

25 ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, . renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous

30 xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma,

35 myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone,

bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

5 ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

10 syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

15 complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding TMP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TMP expression. Such qualitative or quantitative

20 methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TMP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TMP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable

25 incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TMP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the

30 treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TMP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TMP, under conditions suitable for hybridization or amplification. Standard

35 hybridization may be quantified by comparing the values obtained from normal subjects with values

from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

5   Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

10   With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further

15   progression of the cancer.

         Additional diagnostic uses for oligonucleotides designed from the sequences encoding TMP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TMP, or a fragment of a polynucleotide complementary to the polynucleotide encoding TMP, and will be

20   employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

         In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TMP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions,

25   insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TMP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily

30   fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of

35   identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which

assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system

5   (Sequenom, San Diego CA).

Methods which may also be used to quantify the expression of TMP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be

10   accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray

15   can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray·may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the

20   activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TMP, fragments of TMP, or antibodies specific for TMP may be used

25   as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

30   quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the

35   hybridization takes place in high-throughput format, wherein the polynucleotides of the present

invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the
5   case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed
10 molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information
15 from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a
20 toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed
25 gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present
30 invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global
35 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome

67

can be subjected individually to further analysis. Proteome expression patterns, or profiles, are
analyzed by quantifying the number of expressed proteins and their relative abundance under given
conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and
analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is

5   achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by
isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl
sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are
visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent
such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is

10   generally proportional to the level of the protein in the sample. The optical densities of equivalently
positioned protein spots from different samples, for example, from biological samples either treated or
untreated with a test compound or therapeutic agent, are compared to identify any changes in protein
spot density related to the treatment. The proteins in the spots are partially sequenced using, for
example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry.

15   The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of
at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In
some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TMP to quantify the
levels of TMP expression. In one embodiment, the antibodies are used as elements on a microarray,

20   and protein expression levels are quantified by exposing the microarray to the sample and detecting the
levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111;
Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of
methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-
reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

25   Toxicant signatures at the proteome level are also useful for toxicological screening, and should
be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation
between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J.
Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the
analysis of compounds which do not significantly affect the transcript image, but which alter the

30   proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid
degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological
sample containing proteins with the test compound. Proteins that are expressed in the treated biological
sample are separated so that the amount of each protein can be quantified. The amount of each protein

35   is compared to the amount of the corresponding protein in an untreated biological sample. A difference

in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

5   sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

10   Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well

15   known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TMP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be

20   preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1

25   constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for

30   example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TMP on a physical map

35   and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA

associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may

5   reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g.,

10   Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TMP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug

15   screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TMP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT

20   application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TMP, or fragments thereof, and washed. Bound TMP is then detected by methods well known in the art. Purified TMP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

25   In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TMP specifically compete with a test compound for binding TMP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TMP.

In additional embodiments, the nucleotide sequences which encode TMP may be used in any

30   molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are,

35   therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure

in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/244,017, U.S. Ser. No. 60/252,855, U.S. Ser. No. 60/251,825, and U.S. Ser. No. 60/255,085, are hereby expressly incorporated by reference.

5

EXAMPLES

I.   Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized

10   and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

15   Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA

20   purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units

25   5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs

30   were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells

35   including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX

71

DH10B from Life Technologies.

II.   Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least

5   one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

10   Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner

15   (Labsystems Oy, Helsinki, Finland).

III.   Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler

20   (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were

25   carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques

30   disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the

35   GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS,

DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus

5   primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA

10   assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide

15   sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and

20   polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold

25   parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity

30   between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:18-34. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

35   IV.   Identification and Editing of Coding Sequences from Genomic DNA

Putative transmembrane proteins were initially identified by running the Genscan gene

identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a

general-purpose gene identification program which analyzes genomic DNA sequences from a variety of

organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin

5   (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an

assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a

FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for

Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA

sequences encode transmembrane proteins, the encoded polypeptides were analyzed by querying against

10   PFAM models for transmembrane proteins. Potential transmembrane proteins were also identified by

homology to Incyte cDNA sequences that had been annotated as transmembrane proteins. These

selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri

public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison

to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as

15   extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA

coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte

cDNA coverage was available, this information was used to correct or confirm the Genscan predicted

sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding

sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process

20   described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from

edited or unedited Genscan-predicted coding sequences.

V.   Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification

25   program described in Example IV. Partial cDNAs assembled as described in Example III were mapped

to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from

one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory

and dynamic programming to integrate cDNA and genomic information, generating possible splice

variants that were subsequently confirmed, edited, or extended to create a full length sequence.

30   Sequence intervals in which the entire length of the interval was present on more than one sequence in

the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity.

For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals

were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to

be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together

35   by the stitching algorithm in the order that they appear along their parent sequences to generate the

longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public

5   databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST

10   analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the

15   translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to

20   determine whether it contained a complete gene.

VI.   Chromosomal Mapping of TMP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:18-34 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched

25   SEQ ID NO:18-34 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment

30   of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in

35   humans, although this can vary widely due to hot and cold spots of recombination.) The cM

distances are based on genetic markers mapped by Généthon which provide boundaries for radiation
hybrid markers whose sequences were included in each of the clusters. Human genome maps and
other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site
(http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified
5   disease genes map within or in proximity to the intervals indicated above.

     In this manner, SEQ ID NO:22 was mapped to chromosome 11 within the interval from
59.50 to 62.50 centiMorgans and SEQ ID NO:26 was mapped to chromosome 1 within the interval
from 179.2 to 186.4 centiMorgans.

VII.   Analysis of Polynucleotide Expression

10   Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene
and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a
particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995)
supra, ch. 4 and 16.)

     Analogous computer techniques applying BLAST were used to search for identical or related
15   molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is
much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer
search can be modified to determine whether any particular match is categorized as exact or similar.
The basis of the search is the product score, which is defined as:

20

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length
of the sequence match. The product score is a normalized value between 0 and 100, and is calculated
25   as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided
by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by
assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for
every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more
than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The
30   product score represents a balance between fractional overlap and quality in a BLAST alignment. For
example, a product score of 100 is produced only for 100% identity over the entire length of the shorter
of the two sequences being compared. A product score of 70 is produced either by 100% identity and
70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is
produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.
35   Alternatively, polynucleotide sequences encoding TMP are analyzed with respect to the tissue

sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system;

5   embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer,

10   cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TMP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

15   VIII.   Extension of TMP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using

20   OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension

25   was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life

30   Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times;

35   Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN

quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE

and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar,

Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II

5   (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis

on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates,

digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

10   sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For

shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose

gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were

religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs,

15   and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing

media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x

carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham

Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1:

20   94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4

repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN

reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified

using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2,

v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC

25   DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle

sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or

are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides

designed for such extension, and an appropriate genomic library.

30 IX.   Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:18-34 are employed to screen cDNAs, genomic

DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is

specifically described, essentially the same procedure is used with larger nucleotide fragments.

Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National

35   Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine

triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human

5 genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature

10 under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X.   Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

15 photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a

20 substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

25 Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection.

30 After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in

35 detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first

5   strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one

10   with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

15   then dried to completion using a SpeedVAC (Savant Instruments, Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses

20   primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope

25   slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

30   Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

35   Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.

Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate
buffered saline (PBS) (Tropix,, Bedford MA) for 30 minutes at 60°C followed by washes in 0.2%
SDS and distilled water as before.

<u>Hybridization</u>

5   Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and
Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample
mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with
an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly
larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of

10   140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for
about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC,
0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

<u>Detection</u>

       Reporter-labeled hybridization complexes are detected with a microscope equipped with an
15   Innova 70 mixed gas 10 W laser (Coherent, Santa Clara CA) capable of generating spectral lines at
488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is
focused on the array using a 20X microscope objective (Nikon, Melville NY). The slide containing
the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the
objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20

20   micrometers.

       In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.
Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,
Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate
filters positioned between the array and the photomultiplier tubes are used to filter the signals. The

25   emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is
typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,
although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

       The sensitivity of the scans is typically calibrated using the signal intensity generated by a
cDNA control species added to the sample mixture at a known concentration. A specific location on

30   the array contains a complementary DNA sequence, allowing the intensity of the signal at that
location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples
from different sources (e.g., representing test and control cells), each labeled with a different
fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially
expressed, the calibration is done by labeling samples of the calibrating cDNA with the two

35   fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital

(A/D) conversion board (Analog Devices, Norwood MA) installed in an IBM-compatible PC

computer. The digitized data are displayed as an image where the signal intensity is mapped using a

linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high

5   signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping   .

emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot

is centered in each element of the grid. The fluorescence signal within each element is then integrated

10   to obtain a numerical value corresponding to the average intensity of the signal. The software used

for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI.   Complementary Polynucleotides

Sequences complementary to the TMP-encoding sequences, or any parts thereof, are used to

detect, decrease, or inhibit expression of naturally occurring TMP. Although use of oligonucleotides

15   comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with

smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO

4.06 software (National Biosciences) and the coding sequence of TMP. To inhibit transcription, a

complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent

promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is

20   designed to prevent ribosomal binding to the TMP-encoding transcript.

XII.   Expression of TMP

Expression and purification of TMP is achieved using bacterial or virus-based expression

systems. For expression of TMP in bacteria, cDNA is subcloned into an appropriate vector containing

an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription.

25   Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the

T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic

resistant bacteria express TMP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG).

Expression of TMP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with

30   recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as

baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TMP

by either homologous recombination or bacterial-mediated transposition involving transfer plasmid

intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of

cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect

35   cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional

genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

     In most expression systems, TMP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

5   affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from <u>Schistosoma japonicum,</u> enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TMP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification

10   using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, <u>supra</u>, ch. 10 and 16). Purified TMP obtained by these methods can be used directly in the assays shown in Examples XVI and XVII where applicable.

15 **XIII.** **Functional Assays**

     TMP function is assessed by expressing the sequences encoding TMP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

20   contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the

25   recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include

30   changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated

35   Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G.

(1994) Flow Cytometry, Oxford, New York NY.

The influence of TMP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TMP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human

5   immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TMP and other genes of interest can be analyzed by northern analysis or microarray techniques.

10   **XIV.   Production of TMP Specific Antibodies**

TMP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TMP amino acid sequence is analyzed using LASERGENE software

15   (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A

20   peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TMP activity by, for example, binding the peptide or TMP to a substrate, blocking with 1%

25   BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV.   Purification of Naturally Occurring TMP Using Specific Antibodies

Naturally occurring or recombinant TMP is substantially purified by immunoaffinity chromatography using antibodies specific for TMP. An immunoaffinity column is constructed by covalently coupling anti-TMP antibody to an activated chromatographic resin, such as CNBr-activated

30   SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TMP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TMP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TMP binding (e.g.,

35   a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and

TMP is collected.

XVI.   Identification of Molecules Which Interact with TMP

TMP, or biologically active fragments thereof, are labeled with [125]I Bolton-Hunter reagent.

(See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules

5   previously arrayed in the wells of a multi-well plate are incubated with the labeled TMP, washed, and

any wells with labeled TMP complex are assayed. Data obtained using different concentrations of

TMP are used to calculate values for the number, affinity, and association of TMP with the candidate

molecules.

Alternatively, molecules interacting with TMP are analyzed using the yeast two-hybrid

10   system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially

available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

TMP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)

which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions

between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent

15   No. 6,057,101).

XVII.   Demonstration of TMP Activity

Gap Junction Activity of TMP

Gap junction activity of TMP is demonstrated as the ability to induce the formation of

intercellular channels between paired Xenopus laevis oocytes injected with TMP cRNA (Hennemann,

20   supra). One week prior to the experimental injection with TMP cRNA, oocytes are injected with

antisense oligonucleotide to TMP to reduce background. TMP cRNA-injected oocytes are incubated

overnight, stripped of vitelline membranes, and paired for recording of junctional currents by dual cell

voltage clamp. The measured conductances are proportional to gap junction activity of TMP.

Alternatively, an assay for TMP activity measures the ion channel activity of TMP using an

25   electrophysiological assay for ion conductance. TMP can be expressed by transforming a mammalian

cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TMP.

Eukaryotic expression vectors are commercially available, and the techniques to introduce them into

cells are well known to those skilled in the art. A second plasmid which expresses any one of a

number of marker genes, such as ß-galactosidase, is co-transformed into the cells to allow rapid

30   identification of those cells which have taken up and expressed the foreign DNA. The cells are

incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow

expression and accumulation of TMP and ß-galactosidase.

Transformed cells expressing ß-galactosidase are stained blue when a suitable colorimetric

substrate is added to the culture media under conditions that are well known in the art. Stained cells

35   are tested for differences in membrane conductance by electrophysiological techniques that are well

known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or ß-galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TMP will have higher anion or cation conductance relative to control cells. The contribution of TMP to conductance can be confirmed by incubating the cells using antibodies specific for TMP. The

5   antibodies will bind to the extracellular side of TMP, thereby blocking the pore in the ion channel, and the associated conductance.

Transmembrane Protein Activity of TMP

An assay for TMP activity measures the expression of TMP on the cell surface. cDNA encoding TMP is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled

10   with biotin as described (de la Fuente, M. A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using TMP-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of TMP expressed on the cell surface.

An alternative assay for TMP activity is based on a prototypical assay for ligand/receptor-

15   mediated modulation of cell proliferation. This assay measures the amount of newly synthesized DNA in Swiss mouse 3T3 cells expressing TMP. An appropriate mammalian expression vector containing cDNA encoding TMP is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transfected cells are incubated in the presence of [^{3}H]thymidine and varying amounts of TMP ligand. Incorporation of [^{3}H]thymidine into acid-precipitable DNA is measured over an

20   appropriate time interval using a tritium radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold TMP ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of TMP producing a 50% response level, where 100% represents maximal incorporation of [^{3}H]thymidine into acid-precipitable DNA (McKay, I. and Leigh,

25   I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY, p. 73).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

30   Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
6431478	1	6431478CD1	18	6431478CB1
3584654	2	3584654CD1	19	3584654CB1
3737084	3	3737084CD1	20	3737084CB1
71426238	4	71426238CD1	21	71426238CB1
7475123	5	7475123CD1	22	7475123CB1
7481952	6	7481952CD1	23	7481952CB1
382654	7	382654CD1	24	382654CB1
1867351	8	1867351CD1	25	1867351CB1
3323104	9	3323104CD1	26	3323104CB1
4769306	10	4769306CD1	27	4769306CB1
2720058	11	2720058CD1	28	2720058CB1
7481255	12	7481255CD1	29	7481255CB1
1510242	13	1510242CD1	30	1510242CB1
162131	14	162131CD1	31	162131CB1
1837725	15	1837725CD1	32	1837725CB1
3643847	16	3643847CD1	33	3643847CB1
6889872	17	6889872CD1	34	6889872CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	6431478CD1	g310100	4.0E-216	[Rattus norvegicus] developmentally regulated protein
2	3584654CD1	g1054884	0.0	[Rattus norvegicus] prostaglandin F2a receptor regulatory protein Orlicky,D.J. (1996) Negative regulatory activity of a prostaglandin F2a receptor associated protein (FPRP). Prostaglandins Leukot. Essent. Fatty Acids 54, 247-259
3	3737084CD1	g3513451	1.7E-233	[Rattus norvegicus] potassium channel regulator 1 Hoshi,N., et al (1998) KCR1, a membrane protein that facilitates functional expression of non-inactivating K+ currents associates with rat EAG voltage-dependent K+ channels. J. Biol. Chem. 273, 23080-23085
4	71426238CD1	g15990851	1.0E-130	[fl][Homo sapiens] connexin25 (AJ414563)
5	7475123CD1	g3449306	0.0	[Homo sapiens] MEGF7 (multiple EGF-like 7 protein) Nakayama,M. et al., (1998) Genomics 51:27-34
6	7481952CD1	g8926243	1.5E-40	[Mus musculus] low density lipoprotein receptor related protein LRP1B/LRP-DIT
9	33231104CD1	g11414879	0.0	[fl][Homo sapiens] mannosyltransferase Maeda,Y. et al. (2001) PIG-M transfers the first mannose to glycosylphosphatidylinositol on the lumenal side of the ER. EMBO J. 20:250-261

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
11	2720058CD1	g6013381	1.5E-12	TM6P1 (integral membrane protein) [Rattus norvegicus] (Zhang, J. et al. (2000) Biochim. Biophys. Acta 1492:280-284)
12	7481255CD1	g4150939	4.2E-27	[Mus musculus] GSG1
13	1510242CD1	g4263743	5.3E-163	similar to UNC-93; similar to U89424 (PID:g3642687)
14	162131CD1	g13374351	0.0	[f1][Homo sapiens] apolipoprotein L4
15	1837725CD1	g6850311	6.0E-52	Contains similarity to a vacuolar sorting receptor homolog from [Arabidopsis thaliana] gb|U79959
16	3643847CD1	g9944332	2.8E-95	[Mus musculus] Ttyh1 Campbell, H.D. et al. (2000) Genomics 68:89-92
17	6889872CD1	g3191975	1.3E-95	dJ63G5.3 (putative Leucine rich protein) [Homo sapiens]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	6431478CD1	461	S117 S190 S225 S280 S336 S34 T109 T221 T333 T386	N113 N183 N294	Transmembrane domains: M38-T56, V94-I112, Y156-V174, Y195-M216, I231-I248, Y388-M406,	HMMER
					PROTEIN PLACENTAL DIFF33 DEVELOPMENTALLY REGULATED R11H6.2 PD011773:I76-P283, K192-S460	BLAST_PRODOM
					PROTEIN PLACENTAL DIFF33 R11H6.2 DEVELOPMENTALLY REGULATED PD018175:M1-E70	BLAST_PRODOM
2	3584654CD1	879	S150 S158 S244 S288 S349 S46 S574 S580 S598 S623 S66 S72 S747 S758 S772 T303 T402 T472 T505 T506 T570 T757 T820 T866 T97 Y640	N286 N300 N383 N413 N44 N525 N600 N618 N691	Signal cleavage: M1-G21	SPSCAN
					Signal peptide: M1-R25	HMMER
					Transmembrane domain: L833-G851	HMMER
					Immunoglobulin domains: G36-T121, G162-V249, G292-V375, D422-V517, G564-V657, G704-V795	HMMER_PFAM
					PROTEIN PROSTAGLANDIN F2ALPHA RECEPTOR REGULATORY PRECURSOR ASSOCIATED SIGNAL IMMUNOGLOBULIN FOLD PD025644:L9-S814	BLAST_PRODOM
					PROSTAGLANDIN F2ALPHA RECEPTOR REGULATORY PROTEIN PRECURSOR ASSOCIATED SIGNAL IMMUNOGLOBULIN FOLD TRANSMEMBRANE GLYCOPROTEIN PD185779:K815-D879	BLAST_PRODOM
					IMMUNOGLOBULIN DM00001|I39207|25-139:V26-Y132 DM00001|I39207|141-259:V140-E255	BLAST_DOMO
					RGD motif: R703-D705	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	3737084CD1	473	S243 S280 S383 S387 S56 T368 T434 Y417 Y444	N246 N457	Transmembrane domains: M95-Y112, P141-M166, Y257-G277, N396-I419	HMMER
					PROTEIN CHANNEL DIE2 POTASSIUM REGULATOR IONIC C56F8.06C CHROMOSOME I PRECURSOR PD025171: E33-S387, W395-N428, S435-W473	BLAST_PRODOM
					POTASSIUM CHANNEL REGULATOR 1 IONIC CHANNEL PD184319:M1-R32	BLAST_PRODOM
4	71426238CD1	223	T132 T174		Connexin: M1-L201	HMMER_PFAM
					CONNEXINS DM00590	BLAST_DOMO
					Q02739|1-268: M1-P218 P28231|1-269: M1-K208 Q02738|1-262: M1-L206 P08034|1-276: M1-K204	
					GAP JUNCTION CONNEXIN PROTEIN TRANSMEMBRANE ALPHA1 CX43 ALPHA8 ALPHA5 BETA1 PD001135: W3-K108, M16-L201	BLAST_PRODOM
					GAP JUNCTION PROTEIN CONNEXIN TRANSMEMBRANE BETA2 CX26 BETA4 CX31.1 DISEASE PD001118: G119-L201	BLAST_PRODOM
					Connexins proteins. BL00407: S2-V38, A39-F69, P70-Y97, G118-D147, C157-L201	BLIMPS_BLOCKS
					CONNEXIN SIGNATURE PR00206: I20-W44; F51-Q73, L76-A96, L120-Y146, C157-T177, I178-L201	BLIMPS_PRINTS
					transmembrane domain: I23-A39, L120-Y143, T177-F200	HMMER
					Connexins signature 2: C157-P173	MOTIFS
					Connexins signatures connexins_1: L33-T86	PROFILESCAN
					Connexins signatures connexins_2: I136-L193	PROFILESCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7475123CD1	1553	S1007 S1011 S1016 S1065 S1103 S1108 S1117 S1152 S1234 S1313 S1337 S1346 S1416 S1420 S1465 S1479 S149 S1504 S1535 S1548 S177 S21 S287 S366 S390 S404 S430 S496 S529 S540 S562 S583 S61 S72 S761 S804 S816 S843 S848 T1095 T1174 T1211 T1331 T1336 T1508 T384 T389 T453 T461 T475 T485 T623 T629 T727 T793 T880 T907	N1063 N1115 N1409 N1449 N549 N725	signal_cleavage: M1-A20	SPSCAN
					LDL RECEPTOR LIGAND-BINDING REPEAT DM00045 |P01130|37-111: C44-D113, G81-S149, G120-P183	BLAST_DOMO
					LDL RECEPTOR LIGAND-BINDING REPEAT DM00045|P98160|295-336: G81-C122	BLAST_DOMO
					LDL RECEPTOR LIGAND-BINDING REPEAT DM00045|I48623|82-120: C83-C122	BLAST_DOMO
					GLYCOPROTEIN PROTEIN RECEPTOR EGF-LIKE DOMAIN LIPOPROTEIN PRECURSOR SIGNAL TRANSMEMBRANE RECEPTOR RELATED PD149641: R431-D559, D1044-D1171, Q739-D867	BLAST_PRODOM
					LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN PRECURSOR LRP RECEPTOR TRANSMEMBRANE REPEAT ENDOCYTOSIS GLYCOPROTEIN SIGNAL CALCIUM BINDING EGF-LIKE DOMAIN COATED PITS PD126644: V440-I707, I753-Y1040, I1057-C1287	BLAST_PRODOM
					HYPOTHETICAL 294.4 KD PROTEIN HYPOTHETICAL PROTEIN PD126659:C661-I1015	BLAST_PRODOM
					LDL-receptor class A (LDLRA) domain proteins BL01209: C90-E102	BLIMPS_BLOCKS
					LOW DENSITY LIPOPROTEIN PR00261: G81-E102, C42-E63, G158-E179, G201-E222	BLIMPS_PRINTS
					signal peptide: M1-G18	HMMER
					transmembrane domain: I1372-A1390	HMMER
					Low-density lipoprotein receptor repeat: D433-V474 N476-M517 G519-G561 Q563-E604 R605-R645 G741-I782 R784-E825 G827-A869 S871-S911 Y912-A953 G1045-V1086 R1088-R1129 G1131-D1173 R1175-T1220	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5					Low-density lipoprotein receptor domain: P23-L68, P69-P107, Q237-W275, P189-E227, R146-S184, R108-M145,	HMMER_PFAM
6	7481952CD1	1718	S1027 S1031 S1095 S1111 S1165 S1273 S1310 S1463 S1482 S1510 S1517 S1544 S1583 S1686 S200 S214 S240 S274 S28 S366 S553 S565 S580 S812 S866 S877 S938 T1178 T1238 T133 T1359 T1371 T1391 T468 T504 T597 T613 T676 T763 T827	N1235 N126 N1299 N1345 N1545 N1634 N1684 N183 N260 N375 N496 N551 N611 N703 N761 N910 N969 N976	Transmembrane domain: I1642-A1660	HMMER
					Low-density lipoprotein receptor domain: E383-P423, H610-M649, G1043-S1081, S1463-L1502, P1507-N1545, K1546-E1586, R824-T864, E1244-H1283	HMMER-PFAM
					MAM domain: C38-I199, C216-L378, C427-L586, C652-P818, C869-S1027, C1083-T1238, C1291-T1454	HMMER-PFAM
					LDL-receptor class A BL01209: C1485-E1497	BLIMPS-BLOCKS
					MAM domain proteins: BL00740A: C434-W446, BL00740B:R1011-S1031	BLIMPS-BLOCKS
					Low density lipoprotein receptor PR00261: L1476-E1497, R397-E418, K836-E857	BLIMPS-PRINTS
					MAM domain proteins: DM01344|P98072|352-509:E418-D578 DM01344|P28824|595-796:E424-D577 DM01344|A55620|961-1128:C421-I576	BLAST-DOMO
					Cell attachment sequences (RGD): R1011-D1013, R1138-D1140	MOTIFS
7	382654CD1	224	S139 S165 S48 S74 T179 T202		Transmembrane domain: I51-W71	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	1867351CD1	570	S317 S336 S356 S538 S94 T367	N82	Signal peptide: M1-C63	SPScan
					Transmembrane domains: L38-I58, A302-N322, V338-C357, I414-M436, L438-I457, L542-R564	HMMER
					Uncharacterized membrane protein family UPF0013: G43-L204, I264-A426	HMMER-PFAM
					Integral membrane protein PD004336: V344-W462	BLAST-PRODOM
					Integral membrane protein PD149928: I55-P129	BLAST-PRODOM
					Leucine zipper pattern: L537-L558	MOTIFS
9	3323104CD1	423	S141 S203 S3 T263 T417 T48 T73 Y52	N400	Transmembrane domains: L228-Y247, F386-I405	HMMER
					Leucine zipper pattern: L215-L236	MOTIFS
					Intergenic region transmembrane protein: PD040574: W131-I408	BLAST-PRODOM
10	4769306CD1	388	S141 S188 S232 S279 S294 S303 S357 S361 S377 S4 S96 S97 T187 T252 T385 T50	N161 N195 N301 N336 N384	Transmembrane domain: V200-I220	HMMER
					CUB domain: C27-F139	HMMER-PFAM
					Low-density lipoprotein receptor domain: P145-E183	HMMER-PFAM
					LDL-receptor class A BL01209: C166-E178	BLIMPS-BLOCKS
11	2720058CD1	231	S137 T177 T220 T42 Y190	N54 N75	Signal peptide: M1-A14	SPScan
					Transmembrane domains: F6-S26, V95-H115, I196-D215	HMMER
12	7481255CD1	293	S109 S148 S251 S95 T236 T35	N243 N277 N59 N67	Signal peptide: M1-S28	HMMER
					Signal peptide: M1-S22	SPScan
					Transmembrane domains: I126-G144, W156-I184	HMMER
					PMP-22 / EMP / MP20 family BL01221: F88-N101, Y203-T229	BLIMPS-BLOCKS
13	1510242CD1	526	S279 S479 T22 Y125	N180 N201 N378	Transmembrane domains: L146-L172, V217-L235, P333-W351	HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	162131CD1	348	S45 S187 S344 T33 T54 T207 T246 T271 Y39 Y77		transmembrane domain: I149-L175	HMMER
					APOLIPOPROTEIN L PRECURSOR APOL PLASMA LIPID TRANSPORT GLYCOPROTEIN SIGNAL DJ6802.1 PD042084: T9-A347	BLAST_PRODOM
15	1837725CD1	520	S8 S54 S141 T37 T151 T157 T202 T331 T505	N35 N58 N66 N74 N116 N126 N149 N155	signal_cleavage: M1-A25	SPSCAN
					signal peptide: M1-A28	HMMER
					transmembrane domain: M172-G191, F218-L237, W243-Y261, V287-F306, L346-E371	HMMER
16	3643847CD1	534	S35 S40 S124 S133 S338 S512 S519 T143 T170 T339 T418 Y207	N31 N129 N283 N352	signal_cleavage: M1-G57	SPSCAN
					transmembrane domain: L46-F62, L214-G233, L390-C408	HMMER
					TWEETY F42E11.2 PROTEIN PD043235: V20-D429	BLAST_PRODOM MOTIFS
					Cell attachment sequence R164-D166	
17	6889872CD1	820	S92 S160 S164 S217 S289 S363 S367 S530 S568 S636 S639 S672 S725 S737 S741 S747 S757 T56 T87 T338 T378 T437 T483 T493 T625 T766 Y443 Y743 Y785	N54 N80 N85 N117 N205 N247 N329 N371	signal_cleavage: M1-A22	SPSCAN
					signal peptide: M1-A22	HMMER
					transmembrane domain: T394-Y416	HMMER
					Leucine Rich Repeat: T56-G79, N80-S103, S104-S127, R128-P151, S152-A175	HMMER_PFAM

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
18	6431478CB1	2653	1-606, 2109-2306	6798827J1 (COLENOR03)	2076	2653
				6882379J1 (BRAHTDR03)	1679	2342
				7162277F8 (PLACNOR01)	1427	2173
				7757207J1 (SPLNTUE01)	418	1133
				6109577F6 (MCLDTXT03)	870	1464
				7196064F8 (LUNGFER04)	1	735
19	3584654CB1	3531	1-1782	71760758V1	2094	2675
				71760788V1	2684	3350
				71761160V1	2577	3331
				7658349J1 (UTREDME06)	1341	1989
				70815172V1	1481	2064
				71764416V1	2978	3531
				71754337V1	1965	2648
				7183593H1 (BONRFEC01)	1	603
				7709623J1 (PANCNOE02)	895	1480
				6807068F6 (SKIRNOR01)	293	1156
20	3737084CB1	2280	947-997, 1840-2280	70409519D1	1372	1947
				70408370D1	1720	2280
				70400676D1	859	1371
				2310625T6 (NGANNOT01)	1282	1832
				71685057V1	609	1268
				71683716V1	1	748
21	71426238CB1	1104	1-115, 550-624, 198-513	1947421R6 (PITUNOT01)	1	508
22	7475123CB1	4966	1-957, 1328-3163	71433875V1	435	1104
				70879893V1	4038	4682
				8124425H1 (HEAONOC01)	4309	4966
				8253556H1 (BRAHDIT10)	3948	4663
				6891485J1 (BRAITDR03)	2762	3423
				6885135H1 (BRAHTDR03)	1	468
				593337R6 (BRAVUNT02)	3439	3968
				70673072V1	3303	3965

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22				GBI.g7705145_17_12_25_27.regenscan ()	337	3773
23	7481952CB1	5401	4237-4630, 2834-3834, 1-2283, 4757-4851	55153856J1	1947	2415
				GNN.g10120165_000001_006.edit	1	342
				6933885R6 (SINTTMR02)	2518	3182
				6933885H1 (SINTTMR02)	2339	3011
				72434542D1	4614	5401
				55153330J1	3010	3752
				899395H1 (BRSTTUT03)	1589	1891
				55153261H1	311	899
				71876387V1	4346	5105
				71932270V1	3716	4305
				55153293H1	364	1151
				55153285J1	886	1816
				72432008D1	4234	5099
				GNN.g9843576_000009_006	1678	2105
24	382654CB1	1949	1-247, 1882-1949	7629906H1 (BRAFTUE03)	1	517
				6618349J2 (BRAUTDR03)	1282	1949
				6884626H1 (BRAHTDR03)	541	1060
				6952509H1 (BRAITDR02)	1035	1600
				6891983H1 (BRAITDR03)	468	1056
				6891983J1 (BRAITDR03)	1121	1671
25	1867351CB1	2133	987-1049, 1927-2133, 779-832	70571014V1	1520	2133
				70568556V1	1259	1924
				7261108H1 (UTRETMC01)	646	1135
				8004078H1 (MUSCTDC01)	1	704
				70570876V1	1151	1828

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
25				5502 6730J1 (ADMEDNV22)	692	1215
26	3323104CB1	2090	1-34, 415-1111	70275 9T6 (SYNORAT03)	1128	1767
				71220 067V1	585	1119
				70833 651V1	553	1114
				70769 661V1	1061	1692
				81332 88H1 (SCOMDIC01)	1	578
				70771 538V1	1625	2090
27	4769306CB1	1618	1-208, 561-723	70954 880V1	1126	1618
				32696 67H1 (BRAINOT20)	1	244
				70955 603V1	648	1272
				81032 47H1 (MIXDDIE02)	229	797
				70764 27H1 (BRAUTDR04)	49	677
				70953 588V1	795	1368
28	2720058CB1	3269	2536-2564, 1-220, 2785-3269	26575 01T6 (LUNGTUT09)	2502	3056
				71520 348V1	1839	2547
				71524 606V1	1139	1824
				26575 01F6 (LUNGTUT09)	594	1126
				72258 43H1 (LUNGTMC01)	392	983
				71670 642V1	1721	2424
				13051 13F6 (PLACNOT02)	1	571
				19553 63F6 (CONNNOT01)	2570	3058
				13756 65F6 (LUNGNOT10)	2677	3269
				71521 430V1	1054	1714
29	7481255CB1	1227	1-1227	32696 76H1 (BRAINOT20)	5	96
				GBI.g7622436_000021_0 0024.edit	1	1227
				FL74 81255_g8469082_00 0035.g4150939_1_1	99	304
				45385 35F6 (THYRTMT01)	654	1227
30	1510242CB1	2618	1-515, 2475-2618	68231 20J1 (SINTNOR01)	1889	2437
				70367 58F7 (UTRSTMR02)	1	705

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
30				7355340H1 (HEARNON03)	2005	2618
				7930760H1 (COLNDIS02)	1224	1900
				6837255H1 (BRSTNON02)	673	1368
				70366002D1	1430	1904
				3413605H1 (PTHYNOT04)	2456	2618
				71808592V1	806	1433
				6823120H1 (SINTNOR01)	1820	2436
31	162131CB1	2188	1-25, 1173-1633	g1506355	1667	2188
				7243158H2 (PROSTMY01)	1351	1878
				1998635R6 (BRSTUT03)	1632	2172
				70559145V1	454	1196
				70558921V1	1069	1787
				2818229F6 (BRSTNOT14)	1	540
				70558931V1	549	1305
32	1837725CB1	1969	1-400, 1927-1969	70377507D1	1400	1955
				g1378655	1517	1969
				70378490D1	639	1268
33	3643847CB1	3006	796-1137, 2849-3006	6799889J1 (COLENOR03)	1	673
				8227475H1 (BRAUTDR02)	778	1463
				2352377H1 (COLSUCT01)	1763	1959
				1478479H1 (CORPNOT02)	2745	3006
				6993355H1 (BRAQTDR02)	1129	1755
				5964485H1 (BRATNOT05)	1772	2489
				1419930H1 (KIDNNOT09)	2650	2892
				6789167H1 (BRACNOK01)	2091	2801
				6118619H1 (BRAHNON05)	1992	2546
				7436931H1 (ADRETUE02)	594	1202
				72018042V1	1	579
				72018222V1	413	1028
				1992275F6 (CORPNOT02)	1268	1809

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
34	6889872CB1	2884	1-319, 2090-2884, 803-1585, 379-418, 1991-2017	8230901H1 (BRAUTDR02)	873	1610
				8230892H1 (BRAUTDR02)	1	725
				8230892J1 (BRAUTDR02)	676	1356
				7634710H1 (SINTDIE01)	1577	2162
				GNN:g3191973_012	422	2884

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
18	6431478CB1	BRAHTDR03
19	3584654CB1	PROSNOT11
20	3737084CB1	PROSNON01
21	71426238CB1	PITUNOT01
22	7475123CB1	CONUTUT01
23	7481952CB1	SINTTMR02
24	382654CB1	BRAHTDR03
25	1867351CB1	BRAITUT21
26	3323104CB1	LUNGNOT27
27	4769306CB1	BRAMNOT01
28	2720058CB1	LUNGTUT10
29	7481255CB1	THYRTMT01
30	15102242CB1	SINTFER02
31	162131CB1	PANCNOT15
32	18377725CB1	SINDNOT01
33	3643847CB1	CORPNOT02
34	6889872CB1	SINTDIE01

Table 6

Library	Vector	Library Description
BRAHTDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from archaecortex, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAITUT21	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningothelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
BRAMNOT01	pINCY	Library was constructed using RNA isolated from medulla tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
CONUTUT01	pINCY	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.

Table 6 (cont.)

Library	Vector	Library Description
LUNGNOT27	pINCY	Library was constructed using RNA isolated from lung tissue removed from a 17-year-old Hispanic female.
LUNGTUT10	pINCY	Library was constructed using RNA isolated from lung tumor tissue removed from the left upper lobe of a 65-year-old Caucasian female during a segmental lung resection. Pathology indicated a metastatic grade 2 myxoid liposarcoma and a metastatic grade 4 liposarcoma. Patient history included soft tissue cancer, breast cancer, and secondary lung cancer.
PANCNOT15	pINCY	Library was constructed using RNA isolated from diseased pancreatic tissue removed from a 15-year-old Caucasian male during a exploratory laparotomy with distal pancreatectomy and total splenectomy. Pathology indicated islet cell hyperplasia. Family history included prostate cancer and cardiovacular disease.
PITUNOT01	PBLUESCRIPT	Library was constructed using RNA obtained from Clontech (CLON 6584-2, lot 35278). The RNA was isolated from the pituitary glands removed from a pool of 18 male and female Caucasian donors, 16 to 70 years old, who died from trauma.
PROSNON01	PSPORT1	This normalized prostate library was constructed from 4.4 M independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
PROSNOT11	pINCY	Library was constructed using RNA isolated from the prostate tissue of a 28-year-old Caucasian male, who died from a self-inflicted gunshot wound.
SINDNOT01	pINCY	Library was constructed using RNA isolated from duodenum tissue removed from the small intestine of a 16-year-old Caucasian male who died from head trauma. Patient history included a kidney infection.
SINTDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from small intestine tissue removed from a 49-year-old Caucasian female during gastroenterostomy, exploratory laparotomy, and vagotomy. The patient presented with acute stomach ulcer with obstruction, nausea and vomiting, and abnormal weight loss. Patient history included backache, acute stomach ulcer with perforation, and normal delivery. Previous surgeries included adenotonsillectomy and total abdominal hysterectomy. Patient medications included Premarin. Family history included benign hypertension, type II diabetes and congestive heart failure in the father.
SINTFER02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.

Table 6 (cont.)

Library	Vector	Library Description
SINTTMR02	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 59-year-old male. Pathology for the matched tumor tissue indicated multiple (9) carcinoid tumors, grade 1, in the small bowel. The largest tumor was associated with a large mesenteric mass. Multiple convoluted segments of bowel were adhered to the tumor. A single (1 of 13) regional lymph node was positive for malignancy. The peritoneal biopsy indicated focal fat necrosis.
THYRTMT01	pINCY	Library was constructed using RNA isolated from.left thyroid tissue removed from a 56-year-old Caucasian male during a unilateral thyroid lobectomy and fine needle thyroid biopsy. Pathology for the associated tumor tissue indicated medullary carcinoma invading the overlying skeletal muscle. Metastatic medullary carcinoma involved one carotid sheath lymph node (of 9), one left neck lymph node with extra nodular extension, and a central compartment node. A microscopic focus of grade 1 papillary carcinoma was identified within the right lobe of the thyroid lobe. The left thyroid vein biopsy was negative for tumor. Patient history included hyperlipidemia, headache, and atherosclerotic coronary artery disease. Family history included cerebrovascular disease, cardiovasclar disease and bone cancer.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0B-8 or less *Full Length sequences:* Probability value= 1.0B-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta B value=1.06B-6 *Assembled ESTs:* fasta Identity= 95% or greater and Match length=200 bases or greater; fastx B value=1.0B-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0B-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0B-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater, Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

106

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a)   a polypeptide comprising an amino acid sequence selected from the group consisting of
SEQ ID NO:1-17,

b)   a polypeptide comprising a naturally occurring amino acid sequence at least 90%
identical to an amino acid sequence selected from the group consisting of SEQ ID
NO:1-17,

c)   a biologically active fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-17, and

d)   an immunogenic fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-17.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the
group consisting of SEQ ID NO:1-17.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from
the group consisting of SEQ ID NO:18-34.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a
polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a)   culturing a cell under conditions suitable for expression of the polypeptide, wherein
said cell is transformed with a recombinant polynucleotide, and said recombinant
polynucleotide comprises a promoter sequence operably linked to a polynucleotide
encoding the polypeptide of claim 1, and

107

b)   recovering the polypeptide so expressed.

10.   A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-17.

5

11.   An isolated antibody which specifically binds to a polypeptide of claim 1.

12.   An isolated polynucleotide selected from the group consisting of:

a)   a polynucleotide comprising a polynucleotide sequence selected from the group

10   consisting of SEQ ID NO:18-34,

b)   a polynucleotide comprising a naturally occurring polynucleotide sequence at least

         90% identical to a polynucleotide sequence selected from the group consisting of SEQ

         ID NO:18-34,

c)   a polynucleotide complementary to a polynucleotide of a),

15   d)   a polynucleotide complementary to a polynucleotide of b), and

e)   an RNA equivalent of a)-d).

13.   An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

20

14.   A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a)   hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

         comprising a sequence complementary to said target polynucleotide in the sample, and

25   which probe specifically hybridizes to said target polynucleotide, under conditions

         whereby a hybridization complex is formed between said probe and said target

         polynucleotide or fragments thereof, and

b)   detecting the presence or absence of said hybridization complex, and, optionally, if

         present, the amount thereof.

30

15.   A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16.   A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

35   a)   amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and

b)   detecting the presence or absence of said amplified target polynucleotide or fragment
thereof, and, optionally, if present, the amount thereof.

5   17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable
excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence
selected from the group consisting of SEQ ID NO:1-17.

10

19. A method for treating a disease or condition associated with decreased expression of
functional TMP, comprising administering to a patient in need of such treatment the composition of
claim 17.

15   20. A method of screening a compound for effectiveness as an agonist of a polypeptide of
claim 1, the method comprising:

a)   exposing a sample comprising a polypeptide of claim 1 to a compound, and

b)   detecting agonist activity in the sample.

20   21. A composition comprising an agonist compound identified by a method of claim 20 and a
pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of
functional TMP, comprising administering to a patient in need of such treatment a composition of claim
25   21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of
claim 1, the method comprising:

a)   exposing a sample comprising a polypeptide of claim 1 to a compound, and

30   b)   detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and
a pharmaceutically acceptable excipient.

35   25. A method for treating a disease or condition associated with overexpression of functional

TMP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a)   combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b)   detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a)   combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b)   assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c)   comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a)   exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b)   detecting altered expression of the target polynucleotide, and

c)   comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a)   treating a biological sample containing nucleic acids with the test compound,

b)   hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

110

polynucleotide in the biological sample, said target polynucleotide comprising a

polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c)   quantifying the amount of hybridization complex, and

d)   comparing the amount of hybridization complex in the treated biological sample with

5   the amount of hybridization complex in an untreated biological sample, wherein a

difference in the amount of hybridization complex in the treated biological sample is

indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of TMP in a

10   biological sample, the method comprising:

a)   combining the biological sample with an antibody of claim 11, under conditions

suitable for the antibody to bind the polypeptide and form an antibody:polypeptide

complex, and

b)   detecting the complex, wherein the presence of the complex correlates with the

15   presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a)   a chimeric antibody,

b)   a single chain antibody,

20   c)   a Fab fragment,

d)   a $F(ab')_2$ fragment, or

e)   a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

25

33. A method of diagnosing a condition or disease associated with the expression of TMP in a

subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

30

35. A method of diagnosing a condition or disease associated with the expression of TMP in a

subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

35   11, the method comprising:

111

a)   immunizing an animal with a polypeptide consisting of an amino acid sequence selected
        from the group consisting of SEQ ID NO:1-17, or an immunogenic fragment thereof,
        under conditions to elicit an antibody response,

b)   isolating antibodies from said animal, and

c)   screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal
        antibody which binds specifically to a polypeptide comprising an amino acid sequence
        selected from the group consisting of SEQ ID NO:1-17.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim
11, the method comprising:

a)   immunizing an animal with a polypeptide consisting of an amino acid sequence selected
        from the group consisting of SEQ ID NO:1-17, or an immunogenic fragment thereof,
        under conditions to elicit an antibody response,

b)   isolating antibody producing cells from the animal,

c)   fusing the antibody producing cells with immortalized cells to form monoclonal
        antibody-producing hybridoma cells,

d)   culturing the hybridoma cells, and

e)   isolating from the culture monoclonal antibody which binds specifically to a
        polypeptide comprising an amino acid sequence selected from the group consisting of
        SEQ ID NO:1-17.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression
library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant
immunoglobulin library.

112

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 in a sample, the method comprising:

a)   incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

5   b)   detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-17 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the

10   group consisting of SEQ ID NO:1-17 from a sample, the method comprising:

a)   incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b)   separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID

15   NO:1-17.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim

13.

20   47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a)   labeling the polynucleotides of the sample,

b)   contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex,

25   and

c)   quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or

30   polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

35

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide . molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

5

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

10

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

15

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

20

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

25

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

30

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

35

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

5   84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

10

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

15   89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

<110> INCYTE GENOMICS, INC.
     WARREN, Bridget A.
     XU, Yuming
     YUE, Henry
     BATRA, Sajeev
     BURFORD, Neil
     GANDHI, Ameena R.
     WALIA, Narinder K.
     ARVIZU, Chandra
     TANG, Y. Tom
     LU, Dyung Aina M.
     DUGGAN, Brendan M.
     BAUGHN, Mariah R.
     LEE, Ernestine A.
     KHAN, Farrah A.
     NGUYEN, Danniel B.
     AZIMZAI, Yalda
     YAO, Monique G.
     LAL, Preeti G.
     THANGAVELU, Kavitha
     RAMKUMAR, Jayalaxmi
     TRAN, Bao
     DING, Li
     AU-YOUNG, Janice

<120> TRANSMEMBRANE PROTEINS

<130> PF-0836 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/244,017; 60/252,855; 60/251,825; 60/255,085
<151> 2000-10-27; 2000-11-22; 2000-12-07; 2000-12-12

<160> 34
<170> PERL Program

<210> 1
<211> 461
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6431478CD1

<400> 1
Met Ser Ala Gln Cys Cys Ala Gly Gln Leu Ala Cys Cys Cys Gly
 1   5   10   .15
Ser Ala Gly Cys Ser Leu Cys Cys Asp Cys Cys Pro Arg Ile Arg
                20   25   30
Gln Ser Leu Ser Thr Arg Phe Met Tyr Ala Leu Tyr Phe Ile Leu
                35   40   45
Val Val Val Leu Cys Cys Ile Met Met Ser Thr Thr Val Ala His
                50   55   60

```
Lys Met Lys Glu His Ile Pro Phe Phe Glu Asp Met Cys Lys Gly
              65                  70                  75
Ile Lys Ala Gly Asp Thr Cys Glu Lys Leu Val Gly Tyr Ser Ala
              80                  85                  90
Val Tyr Arg Val Cys Phe Gly Met Ala Cys Phe Phe Phe Ile Phe
              95                  100                 105
Cys Leu Leu Thr Leu Lys Ile Asn Asn Ser Lys Ser Cys Arg Ala
 ·            110                 115                 120
His Ile His Asn Gly Phe Trp Phe Phe Lys Leu Leu Leu Leu Gly
              125                 130                 135
Ala Met Cys Ser Gly Ala Phe Phe Ile Pro Asp Gln Asp Thr Phe
              140                 145                 150
Leu Asn Ala Trp Arg Tyr Val Gly Ala Val Gly Gly Phe Leu Phe
              155                 160                 165
Ile Gly Ile Gln Leu Leu Leu Leu Val Glu Phe Ala His Lys Trp
              170                 175                 180
Asn Lys Asn Trp Thr Ala Gly Thr Ala Ser Asn Lys Leu Trp Tyr
              185                 190                 195
Ala Ser Leu Ala Leu Val Thr Leu Ile Met Tyr Ser Ile Ala Thr
              200                 205                 210
Gly Gly Leu Val Leu Met Ala Val Phe Tyr Thr Gln Lys Asp Ser
              215                 220                 225
Cys Met Glu Asn Lys Ile Leu Leu Gly Val Asn Gly Gly Leu Cys
              230                 235                 240
Leu Leu Ile Ser Leu Val Ala Ile Ser Pro Trp Val Gln Asn Arg
              245                 250                 255
Gln Pro His Ser Gly Leu Leu Gln Ser Gly Val Ile Ser Cys Tyr
              260                 265                 270
Val Thr Tyr Leu Thr Phe Ser Ala Leu Ser Ser Lys Pro Ala Glu
              275                 280                 285
Val Val Leu Asp Glu His Gly Lys Asn Val Thr Ile Cys Val Pro
              290                 295                 300
Asp Phe Gly Gln Asp Leu Tyr Arg Asp Glu Asn Leu Val Thr Ile
              305                 310                 315
Leu Gly Thr Ser Leu Leu Ile Gly Cys Ile Leu Tyr Ser Cys Leu
              320                 325                 330
Thr Ser Thr Thr Arg Ser Ser Ser Asp Ala Leu Gln Gly Arg Tyr
              335                 340                 345
Ala Ala Pro Glu Leu Glu Ile Ala Arg Cys Cys Phe Cys Phe Ser
              350                 355                 360
Pro Gly Gly Glu Asp Thr Glu Glu Gln Gln Pro Gly Lys Glu Gly
              365                 370                 375
Pro Arg Val Ile Tyr Asp Glu Lys Lys Gly Thr Val Tyr Ile Tyr
              380                 385                 390
Ser Tyr Phe His Phe Val Phe Phe Leu Ala Ser Leu Tyr Val Met
              395                 400                 405
Met Thr Val Thr Asn Trp Phe Asn Tyr Glu Ser Ala Asn Ile Glu
              410                 415                 420
Ser Phe Phe Ser Gly Ser Trp Ser Ile Phe Trp Val Lys Met Ala
              425                 430                 435
Ser Cys Trp Ile Cys Val Leu Leu Tyr Leu Cys Thr Leu Val Ala
              440                 445                 450
Pro Leu Cys Cys Pro Thr Arg Glu Phe Ser Val
              455                 460
```

<210> 2

<211> 879
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3584654CD1

<400> 2
Met Gly Arg Leu Ala Ser Arg Pro Leu Leu Leu Ala Leu Leu Ser
 1   5   10   15
Leu Ala Leu Cys Arg Gly Arg Val Val Arg Val Pro Thr Ala Thr
                20   25   30
Leu Val Arg Val Val Gly Thr Glu Leu Val Ile Pro Cys Asn Val
                35   40   45
Ser Asp Tyr Asp Gly Pro Ser Glu Gln Asn Phe Asp Trp Ser Phe
                50   55   60
Ser Ser Leu Gly Ser Ser Phe Val Glu Leu Ala Ser Thr Trp Glu
                65   70   75
Val Gly Phe Pro Ala Gln Leu Tyr Gln Glu Arg Leu Gln Arg Gly
                80   85   90
Glu Ile Leu Leu Arg Arg Thr Ala Asn Asp Ala Val Glu Leu His
                95   100   105
Ile Lys Asn Val Gln Pro Ser Asp Gln Gly His Tyr Lys Cys Ser
                110   115   120
Thr Pro Ser Thr Asp Ala Thr Val Gln Gly Asn Tyr Glu Asp Thr
                125   130   135
Val Gln Val Lys Val Leu Ala Asp Ser Leu His Val Gly Pro Ser
                140   145   150
Ala Arg Pro Pro Pro Ser Leu Ser Leu Arg Glu Gly Glu Pro Phe
                155   160   165
Glu Leu Arg Cys Thr Ala Ala Ser Ala Ser Pro Leu His Thr His
                170   175   180
Leu Ala Leu Leu Trp Glu Val His Arg Gly Pro Ala Arg Arg Ser
                185   190   195
Val Leu Ala Leu Thr His Glu Gly Arg Phe His Pro Gly Leu Gly
                200   205   210
Tyr Glu Gln Arg Tyr His Ser Gly Asp Val Arg Leu Asp Thr Val
                215   220   225
Gly Ser Asp Ala Tyr Arg Leu Ser Val Ser Arg Ala Leu Ser Ala
                230   235   240
Asp Gln Gly Ser Tyr Arg Cys Ile Val Ser Glu Trp Ile Ala Glu
                245   250   255
Gln Gly Asn Trp Gln Glu Ile Gln Glu Lys Ala Val Glu Val Ala
                260   265   270
Thr Val Val Ile Gln Pro Thr Val Leu Arg Ala Ala Val Pro Lys
                275   280   285
Asn Val Ser Val Ala Glu Gly Lys Glu Leu Asp Leu Thr Cys Asn
                290   295   300
Ile Thr Thr Asp Arg Ala Asp Asp Val Arg Pro Glu Val Thr Trp
                305   310   315
Ser Phe Ser Arg Met Pro Asp Ser Thr Leu Pro Gly Ser Arg Val
                320   325   330
Leu Ala Arg Leu Asp Arg Asp Ser Leu Val His Ser Ser Pro His
                335   340   345
Val Ala Leu Ser His Val Asp Ala Arg Ser Tyr His Leu Leu Val

```
                    350                     355                     360
Arg Asp Val Ser Lys Glu Asn Ser Gly Tyr Tyr Tyr Cys His Val
                    365                     370                     375
Ser Leu Trp Ala Pro Gly His Asn Arg Ser Trp His Lys Val Ala
                    380                     385                     390
Glu Ala Val Ser Ser Pro Ala Gly Val Gly Val Thr Trp Leu Glu
                    395                     400                     405
Pro Asp Tyr Gln Val Tyr Leu Asn Ala Ser Lys Val Pro Gly Phe
                    410                     415                     420
Ala Asp Asp Pro Thr Glu Leu Ala Cys Arg Val Val Asp Thr Lys
                    425                     430                     435
Ser Gly Glu Ala Asn Val Arg Phe Thr Val Ser Trp Tyr Tyr Arg
                    440                     445                     450
Met Asn Arg Arg Ser Asp Asn Val Val Thr Ser Glu Leu Leu Ala
                    455                     460                     465
Val Met Asp Gly Asp Trp Thr Leu Lys Tyr Gly Glu Arg Ser Lys
                    470                     475                     480
Gln Arg Ala Gln Asp Gly Asp Phe Ile Phe Ser Lys Glu His Thr
                    485                     490                     495
Asp Thr Phe Asn Phe Arg Ile Gln Arg Thr Thr Glu Glu Asp Arg
                    500                     505                     510
Gly Asn Tyr Tyr Cys Val Val Ser Ala Trp Thr Lys Gln Arg Asn
                    515                     520                     525
Asn Ser Trp Val Lys Ser Lys Asp Val Phe Ser Lys Pro Val Asn
                    530                     535                     540
Ile Phe Trp Ala Leu Glu Asp Ser Val Leu Val Val Lys Ala Arg
                    545                     550                     555
Gln Pro Lys Pro Phe Phe Ala Ala Gly Asn Thr Phe Glu Met Thr
                    560                     565                     570
Cys Lys Val Ser Ser Lys Asn Ile Lys Ser Pro Arg Tyr Ser Val
                    575                     580                     585
Leu Ile Met Ala Glu Lys Pro Val Gly Asp Leu Ser Ser Pro Asn
                    590                     595                     600
Glu Thr Lys Tyr Ile Ile Ser Leu Asp Gln Asp Ser Val Val Lys
                    605                     610                     615
Leu Glu Asn Trp Thr Asp Ala Ser Arg Val Asp Gly Val Val Leu
                    620                     625                     630
Glu Lys Val Gln Glu Asp Glu Phe Arg Tyr Arg Met Tyr Gln Thr
                    635                     640                     645
Gln Val Ser Asp Ala Gly Leu Tyr Arg Cys Met Val Thr Ala Trp
                    650                     655                     660
Ser Pro Val Arg Gly Ser Leu Trp Arg Glu Ala Ala Thr Ser Leu
                    665                     670                     675
Ser Asn Pro Ile Glu Ile Asp Phe Gln Thr Ser Gly Pro Ile Phe
                    680                     685                     690
Asn Ala Ser Val His Ser Asp Thr Pro Ser Val Ile Arg Gly Asp
                    695                     700                     705
Leu Ile Lys Leu Phe Cys Ile Ile Thr Val Glu Gly Ala Ala Leu
                    710                     715                     720
Asp Pro Asp Asp Met Ala Phe Asp Val Ser Trp Phe Ala Val His
                    725                     730                     735
Ser Phe Gly Leu Asp Lys Ala Pro Val Leu Leu Ser Ser Leu Asp
                    740                     745                     750
Arg Lys Gly Ile Val Thr Thr Ser Arg Arg Asp Trp Lys Ser Asp
                    755                     760                     765
Leu Ser Leu Glu Arg Val Ser Val Leu Glu Phe Leu Leu Gln Val
```

```
                    770                  775                  780
His Gly Ser Glu Asp Gln Asp Phe Gly Asn Tyr Tyr Cys Ser Val
                    785                  790                  795
Thr Pro Trp Val Lys Ser Pro Thr Gly Ser Trp Gln Lys Glu Ala
                    800                  805                  810
Glu Ile His Ser Lys Pro Val Phe Ile Thr Val Lys Met Asp Val
                    815                  820                  825
Leu Asn Ala Phe Lys Tyr Pro Leu Leu Ile Gly Val Gly Leu Ser
                    830                  835                  840
Thr Val Ile Gly Leu Leu Ser Cys Leu Ile Gly Tyr Cys Ser Ser
                    845                  850                  855
His Trp Cys Cys Lys Lys Glu Val Gln Glu Thr Arg Arg Glu Arg
                    860                  865                  870
Arg Arg Leu Met Ser Met Glu Met Asp
                    875
```

```
<210> 3
<211> 473
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3737084CD1

<400> 3
Met Ala Gln Leu Glu Gly Tyr Tyr Phe Ser Ala Ala Leu Ser Cys
1                   5                   10                  15
Thr Phe Leu Val Ser Cys Leu Leu Phe Ser Ala Phe Ser Arg Ala
                    20                  25        ·         30
Leu Arg Glu Pro Tyr Met Asp Glu Ile Phe His Leu Pro Gln Ala
                    35                  40                  45
Gln Arg Tyr Cys Glu Gly His Phe Ser Leu Ser Gln Trp Asp Pro
                    50  ·               .55                 60
Met Ile Thr Thr Leu Pro Gly Leu Tyr Leu Val Ser Ile Gly Val
                    65                  70                  75
Ile Lys Pro Ala Ile Trp Ile Phe Gly Trp Ser Glu His Val Val
                    80                  85                  90
Cys Ser Ile Gly Met Leu Arg Phe Val Asn Leu Leu Phe Ser Val
                    95                  100                 105
Gly Asn Phe Tyr Leu Leu Tyr Leu Leu Phe Cys Lys Val Gln Pro
                    110                 115                 120
Arg Asn Lys Ala Ala Ser Ser Ile Gln Arg Val Leu Ser Thr Leu
                    125                 130                 135
Thr Leu Ala Val Phe Pro Thr Leu Tyr Phe Phe Asn Phe Leu Tyr
                    140                 145                 150
Tyr Thr Glu Ala Gly Ser Met Phe Phe Thr Leu Phe Ala Tyr Leu
                    155                 160                 165
Met Cys Leu Tyr Gly Asn His Lys Thr Ser Ala Phe Leu Gly Phe
                    170                 175                 180
Cys Gly Phe Met Phe Arg Gln Thr Asn Ile Ile Trp Ala Val Phe
                    185                 190                 195
Cys Ala Gly Asn Val Ile Ala Gln Lys Leu Thr Glu Ala Trp Lys
                    200                 205                 210
Thr Glu Leu Gln Lys Lys Glu Asp Arg Leu Pro Pro Ile Lys Gly
                    215                 220                 225
```

```
Pro Phe Ala Glu Phe Arg Lys Ile Leu Gln Phe Leu Leu Ala Tyr
            230                 235                 240
Ser Met Ser Phe Lys Asn Leu Ser Met Leu Leu Leu Thr Trp
            245                 250                 255
Pro Tyr Ile Leu Leu Gly Phe Leu Phe Cys Ala Phe Val Val Val
            260                 265                 270
Asn Gly Gly Ile Val Ile Gly Asp Arg Ser Ser His Glu Ala Cys
            275                 280                 285
Leu His Phe Pro Gln Leu Phe Tyr Phe Phe Ser Phe Thr Leu Phe
            290                 295                 300
Phe Ser Phe Pro His Leu Leu Ser Pro Ser Lys Ile Lys Thr Phe
            305                 310                 315
Leu Ser Leu Val Trp Lys Arg Arg Ile Leu Phe Phe Val Val Thr
            320                 325                 330
Leu Val Ser Val Phe Leu Val Trp Lys Phe Thr Tyr Ala His Lys
            335                 340                 345
Tyr Leu Leu Ala Asp Asn Arg His Tyr Thr Phe Tyr Val Trp Lys
            350                 355                 360
Arg Val Phe Gln Arg Tyr Glu Thr Val Lys Tyr Leu Leu Val Pro
            365                 370                 375
Ala Tyr Ile Phe Ala Gly Trp Ser Ile Ala Asp Ser Leu Lys Ser
            380                 385                 390
Lys Ser Ile Phe Trp Asn Leu Met Phe Phe Ile Cys Leu Phe Thr
            395                 400                 405
Val Ile Val Pro Gln Lys Leu Leu Glu Phe Arg Tyr Phe Ile Leu
            410                 415                 420
Pro Tyr Val Ile Tyr Arg Leu Asn Ile Pro Leu Pro Pro Thr Ser
            425                 430                 435
Arg Leu Ile Cys Glu Leu Ser Cys Tyr Ala Val Val Asn Phe Ile
            440                 445                 450
Thr Phe Phe Ile Phe Leu Asn Lys Thr Phe Gln Trp Pro Asn Ser
            455                 460                 465
Gln Asp Ile Gln Arg Phe Met Trp
            470
```

```
<210> 4
<211> 223
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71426238CD1

<400> 4
Met Ser Trp Met Phe Leu Arg Asp Leu Leu Ser Gly Val Asn Lys
 1               5                  10                  15
Tyr Ser Thr Gly Ile Gly Trp Ile Trp Leu Ala Val Val Phe Val
            20                  25                  30
Phe Arg Leu Leu Val Tyr Met Val Ala Ala Glu His Val Trp Lys
            35                  40                  45
Asp Glu Gln Lys Glu Phe Glu Cys Asn Ser Arg Gln Pro Gly Cys
            50                  55                  60
Lys Asn Val Cys Phe Asp Asp Phe Phe Pro Ile Ser Gln Val Arg
            65                  70                  75
Leu Trp Ala Leu Gln Leu Ile Met Val Ser Thr Pro Ser Leu Leu
```

```
                      80                    85                    90
Val Val Leu His Val Ala Tyr His Glu Gly Arg Glu Lys Arg His
                      95                   100                   105
Arg Lys Lys Leu Tyr Val Ser Pro Gly Thr Met Asp Gly Gly Leu
                     110                   115                   120
Trp Tyr Ala Tyr Leu Ile Ser Leu Ile Val Lys Thr Gly Phe Glu
                     125                   130                   135
Ile Gly Phe Leu Val Leu Phe Tyr Lys Leu Tyr Asp Gly Phe Ser
                     140                   145                   150
Val Pro Tyr Leu Ile Lys Cys Asp Leu Lys Pro Cys Pro Asn Thr
                     155                   160                   165
Val Asp Cys Phe Ile Ser Lys Pro Thr Glu Lys Thr Ile Phe Ile
                     170                   175                   180
Leu Phe Leu Val Ile Thr Ser Cys Leu Cys Ile Val Leu Asn Phe
                     185                   190                   195
Ile Glu Leu Ser Phe Leu Val Leu Lys Cys Leu Ile Lys Cys Cys
                     200                   205                   210
Leu Gln Lys Tyr Leu Lys Lys Pro Gln Val Leu Ser Val
                     215                   220
```

```
<210> 5
<211> 1553
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475123CD1

<400> 5
Met Arg Arg Gln Trp Gly Ala Leu Leu Leu Gly Ala Leu Leu Cys
  1                   5                    10                   15
Ala His Gly Leu Ala Ser Ser Pro Glu Cys Ala Cys Gly Arg Ser
                     20                    25                   30
His Phe Thr Cys Ala Val Ser Ala Leu Gly Glu Cys Thr Cys Ile
                     35                    40                   45
Pro Ala Gln Trp Gln Cys Asp Gly Asp Asn Asp Cys Gly Asp His
                     50                    55                   60
Ser Asp Glu Asp Gly Cys Ile Leu Pro Thr Cys Ser Pro Leu Asp
                     65                    70                   75
Phe His Cys Asp Asn Gly Lys Cys Ile Arg Arg Ser Trp Val Cys
                     80                    85                   90
Asp Gly Asp Asn Asp Cys Glu Asp Asp Ser Asp Glu Gln Asp Cys
                     95                   100                   105
Pro Pro Arg Glu Cys Glu Glu Asp Glu Phe Pro Cys Gln Asn Gly
                     110                   115                   120
Tyr Cys Ile Arg Ser Leu Trp His Cys Asp Gly Asp Asn Asp Cys
                     125                   130                   135
Gly Asp Asn Ser Asp Glu Gln Cys Asp Met Arg Lys Cys Ser Asp
                     140                   145                   150
Lys Glu Phe Arg Cys Ser Asp Gly Ser Cys Ile Ala Glu His Trp
                     155                   160                   165
Tyr Cys Asp Gly Asp Thr Asp Cys Lys Asp Gly Ser Asp Glu Glu
                     170                   175                   180
Asn Cys Pro Ser Ala Val Pro Ala Pro Pro Cys Asn Leu Glu Glu
                     185                   190                   195
```

```
Phe Gln Cys Ala Tyr Gly Arg Cys Ile Leu Asp Ile Tyr His Cys
            200                 205                 210
Asp Gly Asp Asp Asp Cys Gly Asp Trp Ser Asp Glu Ser Asp Cys
            215                 220                 225
Cys Glu Tyr Ser Gly Gln Leu Gly Ala Ser His Gln Pro Cys Arg
            230                 235                 240
Ser Gly Glu Phe Met Cys Asp Ser Gly Leu Cys Ile Asn Ala Gly
            245                 250                 255
Trp Arg Cys Asp Gly Asp Ala Asp Cys Asp Asp Gln Ser Asp Glu
            260                 265                 270
Arg Asn Cys Asn Trp Gln Thr Lys Ser Ile Gln Arg Val Asp Lys
            275                 280                 285
Tyr Ser Gly Arg Asn Lys Glu Thr Val Leu Ala Asn Val Glu Gly
            290                 295                 300
Leu Met Asp Ile Ile Val Val Ser Pro Gln Arg Gln Thr Gly Thr
            305                 310                 315
Asn Ala Cys Gly Val Asn Asn Gly Gly Cys Thr His Leu Cys Phe
            320                 325                 330
Ala Arg Ala Ser Asp Phe Val Cys Ala Cys Pro Asp Glu Pro Asp
            335                 340                 345
Ser Arg Pro Cys Ser Leu Val Pro Gly Leu Val Pro Pro Ala Pro
            350                 355                 360
Arg Ala Thr Gly Met Ser Glu Lys Ser Pro Val Leu Pro Asn Thr
            365                 370                 375
Pro Pro Thr Thr Leu Tyr Ser Ser Thr Thr Arg Thr Arg Thr Ser
            380                 385                 390
Leu Glu Glu Val Glu Gly Arg Met Asp Ile Arg Arg Ile Ser Phe
            395                 400                 405
Asp Thr Glu Asp Leu Ser Asp Asp Val Ile Pro Leu Ala Asp Val
            410                 415                 420
Arg Ser Ala Val Ala Leu Asp Trp Asp Ser Arg Asp Asp His Val
            425                 430                 435
Tyr Trp Thr Asp Val Ser Thr Asp Thr Ile Ser Arg Ala Lys Trp
            440                 445                 450
Asp Gly Thr Gly Gln Glu Val Val Val Asp Thr Ser Leu Glu Ser
            455                 460                 465
Pro Ala Gly Leu Ala Ile Asp Trp Val Thr Asn Lys Leu Tyr Trp
            470                 475                 480
Thr Asp Ala Gly Thr Asp Arg Ile Glu Val Ala Asn Thr Asp Gly
            485                 490                 495
Ser Met Arg Thr Val Leu Ile Trp Glu Asn Leu Asp Arg Pro Arg
            500                 505                 510
Asp Ile Val Val Glu Pro Met Gly Gly Tyr Met Tyr Trp Thr Asp
            515                 520                 525
Trp Gly Ala Ser Pro Lys Ile Glu Arg Ala Gly Met Asp Ala Ser
            530                 535                 540
Gly Arg Gln Val Ile Ile Ser Ser Asn Leu Thr Trp Pro Asn Gly
            545                 550                 555
Leu Ala Ile Asp Tyr Gly Ser Gln Arg Leu Tyr Trp Ala Asp Ala
            560                 565                 570
Gly Met Lys Thr Ile Glu Phe Ala Gly Leu Asp Gly Ser Lys Arg
            575                 580                 585
Lys Val Leu Ile Gly Ser Gln Leu Pro His Pro Phe Gly Leu Thr
            590                 595                 600
Leu Tyr Gly Glu Arg Ile Tyr Trp Thr Asp Trp Gln Thr Lys Ser
            605                 610                 615
```

```
Ile Gln Ser Ala Asp Arg Leu Thr Gly Leu Asp Arg Glu Thr Leu
            620                 625                 630
Gln Glu Asn Leu Glu Asn Leu Met Asp Ile His Val Phe His Arg
            635                 640                 645
Arg Arg Pro Pro Val Ser Thr Pro Cys Ala Met Glu Asn Gly Gly
            650                 655                 660
Cys Ser His Leu Cys Leu Arg Ser Pro Asn Pro Ser Gly Phe Ser
            665                 670                 675
Cys Thr Cys Pro Thr Gly Ile Asn Leu Leu Ser Asp Gly Lys Thr
            680                 685                 690
Cys Ser Pro Gly Met Asn Ser Phe Leu Ile Phe Ala Arg Arg Ile
            695                 700                 705
Asp Ile Arg Met Val Ser Leu Asp Ile Pro Tyr Phe Ala Asp Val
            710                 715                 720
Val Val Pro Ile Asn Ile Thr Met Lys Asn Thr Ile Ala Ile Gly
            725                 730                 735
Val Asp Pro Gln Glu Gly Lys Val Tyr Trp Ser Asp Ser Thr Leu
            740                 745                 750
His Arg Ile Ser Arg Ala Asn Leu Asp Gly Ser Gln His Glu Asp
            755                 760                 765
Ile Ile Thr Thr Gly Leu Gln Thr Thr Asp Gly Leu Ala Val Asp
            770                 775                 780
Ala Ile Gly Arg Lys Val Tyr Trp Thr Asp Thr Gly Thr Asn Arg
            785                 790                 795
Ile Glu Val Gly Asn Leu Asp Gly Ser Met Arg Lys Val Leu Val
            800                 805                 810
Trp Gln Asn Leu Asp Ser Pro Arg Ala Ile Val Leu Tyr His Glu
            815                 820                 825
Met Gly Phe Met Tyr Trp Thr Asp Trp Gly Glu Asn Ala Lys Leu
            830                 835                 840
Glu Arg Ser Gly Met Asp Gly Ser Asp Arg Ala Val Leu Ile Asn
            845                 850                 855
Asn Asn Leu Gly Trp Pro Asn Gly Leu Thr Val Asp Lys Ala Ser
            860                 865                 870
Ser Gln Leu Leu Trp Ala Asp Ala His Thr Glu Arg Ile Glu Ala
            875                 880                 885
Ala Asp Leu Asn Gly Ala Asn Arg His Thr Leu Val Ser Pro Val
            890                 895                 900
Gln His Pro Tyr Gly Leu Thr Leu Leu Asp Ser Tyr Ile Tyr Trp
            905                 910                 915
Thr Asp Trp Gln Thr Arg Ser Ile His Arg Ala Asp Lys Gly Thr
            920                 925                 930
Gly Ser Asn Val Ile Leu Val Arg Ser Asn Leu Pro Gly Leu Met
            935                 940                 945
Asp Met Gln Ala Val Asp Arg Ala Gln Pro Leu Gly Phe Asn Lys
            950                 955                 960
Cys Gly Ser Arg Asn Gly Gly Cys Ser His Leu Cys Leu Pro Arg
            965                 970                 975
Pro Ser Gly Phe Ser Cys Ala Cys Pro Thr Gly Ile Gln Leu Lys
            980                 985                 990
Gly Asp Gly Lys Thr Cys Asp Pro Ser Pro Glu Thr Tyr Leu Leu
            995                 1000                1005
Phe Ser Ser Arg Gly Ser Ile Arg Arg Ile Ser Leu Asp Thr Ser
            1010                1015                1020
Asp His Thr Asp Val His Val Pro Val Pro Glu Leu Asn Asn Val
            1025                1030                1035
```

Ile Ser Leu Asp Tyr Asp Ser Val Asp Gly Lys Val Tyr Tyr Thr
              1040   1045   1050
Asp Val Phe Leu Asp Val Ile Arg Arg Ala Asp Leu Asn Gly Ser
              1055   1060   1065
Asn Met Glu Thr Val Ile Gly Arg Gly Leu Lys Thr Thr Asp Gly
              1070   1075   1080
Leu Ala Val Asp Trp Val Ala Arg Asn Leu Tyr Trp Thr Asp Thr
              1085   1090   1095
Gly Arg Asn Thr Ile Glu Ala Ser Arg Leu Asp Gly Ser Cys Arg
              1100   1105   1110
Lys Val Leu Ile Asn Asn Ser Leu Asp Glu Pro Arg Ala Ile Ala
              1115   1120   1125
Val Phe Pro Arg Lys Gly Tyr Leu Phe Trp Thr Asp Trp Gly His
              1130   1135   1140
Ile Ala Lys Ile Glu Arg Ala Asn Leu Asp Gly Ser Glu Arg Lys
              1145   1150   1155
Val Leu Ile Asn Thr Asp Leu Gly Trp Pro Asn Gly Leu Thr Leu
              1160   1165   1170
Asp Tyr Asp Thr Arg Arg Ile Tyr Trp Val Asp Ala His Leu Asp
              1175   1180   1185
Arg Ile Glu Ser Ala Asp Leu Asn Gly Lys Leu Arg Gln Val Leu
              1190   1195   1200
Val Ser His Val Ser His Pro Phe Ala Leu Thr Gln Gln Asp Arg
              1205   1210   1215
Trp Ile Tyr Trp Thr Asp Trp Gln Thr Lys Ser Ile Gln Arg Val
              1220   1225   1230
Asp Lys Tyr Ser Gly Arg Asn Lys Glu Thr Val Leu Ala Asn Val
              1235   1240   1245
Glu Gly Leu Met Asp Ile Ile Val Val Ser Pro Gln Arg Gln Thr
              1250   1255   1260
Gly Thr Asn Ala Cys Gly Val Asn Asn Gly Gly Cys Thr His Leu
              1265   1270   1275
Cys Phe Ala Arg Ala Ser Asp Phe Val Cys Ala Cys Pro Asp Glu
              1280   1285   1290
Pro Asp Ser Gln Pro Cys Ser Leu Val Pro Gly Leu Val Pro Pro
              1295   1300   1305
Ala Pro Arg Ala Thr Gly Met Ser Glu Lys Ser Pro Val Leu Pro
              1310   1315   1320
Asn Thr Pro Pro Thr Thr Leu Tyr Ser Ser Thr Thr Arg Thr Arg
              1325   1330   1335
Thr Ser Leu Glu Glu Val Glu Gly Arg Cys Ser Glu Arg Asp Ala
              1340   1345   1350
Arg Leu Gly Leu Cys Ala Arg Ser Asn Asp Ala Val Pro Ala Ala
              1355   1360   1365
Pro Gly Glu Gly Leu His Ile Ser Tyr Ala Ile Gly Gly Leu Leu
              1370   1375   1380
Ser Ile Leu Leu Ile Leu Val Val Ile Ala Ala Leu Met Leu Tyr
              1385   1390   1395
Arg His Lys Lys Ser Lys Phe Thr Asp Pro Gly Met Gly Asn Leu
              1400   1405   1410
Thr Tyr Ser Asn Pro Ser Tyr Arg Thr Ser Thr Gln Glu Val Lys
              1415   1420   1425
Ile Glu Ala Ile Pro Lys Pro Ala Met Tyr Asn Gln Leu Cys Tyr
              1430   1435   1440
Lys Lys Glu Gly Gly Pro Asp His Asn Tyr Thr Lys Glu Lys Ile
              1445   1450   1455

```
Lys Ile Val Glu Gly Ile Cys Leu Leu Ser Gly Asp Asp Ala Glu
              1460              1465              1470
Trp Asp Asp Leu Lys Gln Leu Arg Ser Ser Arg Gly Gly Leu Leu
              1475              1480              1485
Arg Asp His Val Cys Met Lys Thr Asp Thr Val Ser Ile Gln Ala
              1490              1495              1500
Ser Ser Gly Ser Leu Asp Asp Thr Glu Thr Glu Gln Leu Leu Gln
              1505              1510              1515
Glu Glu Gln Ser Glu Cys Ser Ser Val His Thr Ala Ala Thr Pro
              1520              1525              1530
Glu Arg Arg Gly Ser Leu Pro Asp Thr Gly Trp Lys His Glu Arg
              1535              1540              1545
Lys Leu Ser Ser Glu Ser Gln Val
              1550
```

<210> 6
<211> 1718
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481952CD1

<400> 6

```
Met Asp Gln Ser Ile Ser Ile Thr Trp Glu Leu Ser Gly Asn Ala
 1               5                10                15
Glu Pro Gln Ala Leu Ala Gln Pro Tyr Arg Thr Lys Ser Tyr Met
                20                25                30
Glu Gln Ala Lys His Leu Thr Cys Asp Phe Glu Ser Gly Phe Cys
                35                40                45
Gly Trp Glu Pro Phe Leu Thr Glu Asp Ser His Trp Lys Leu Met
                50                55                60
Lys Gly Leu Asn Asn Gly Glu His His Phe Pro Ala Ala Asp His
                65                70                75
Thr Ala Asn Ile Asn His Gly Ser Phe Ile Tyr Leu Glu Ala Gln
                80                85                90
Arg Ser Pro Gly Val Ala Lys Leu Gly Ser Pro Val Leu Thr Lys
                95               100               105
Leu Leu Thr Ala Ser Thr Pro Cys Gln Val Gln Phe Trp Tyr His
                110               115               120
Leu Ser Gln His Ser Asn Leu Ser Val Phe Thr Arg Thr Ser Leu
                125               130               135
Asp Gly Asn Leu Gln Lys Gln Gly Lys Ile Ile Arg Phe Ser Glu
                140               145               150
Ser Gln Trp Ser His Ala Lys Ile Asp Leu Ile Ala Glu Ala Gly
                155               160               165
Glu Ser Thr Leu Pro Phe Gln Leu Ile Leu Glu Ala Thr Val Leu
                170               175               180
Ser Ser Asn Ala Thr Val Ala Leu Asp Asp Ile Ser Val Ser Gln
                185               190               195
Glu Cys Glu Ile Ser Tyr Lys Ser Leu Pro Arg Thr Ser Thr Gln
                200               205               210
Ser Lys Phe Ser Lys Cys Asp Phe Glu Ala Asn Ser Cys Asp Trp
                215               220               225
Phe Glu Val Ile Ser Gly Asp His Phe Asp Trp Ile Arg Ser Ser
```

```
                    230                    235                    240
Gln Ser Glu Leu Ser Ala Asp Phe Glu His Gln Ala Pro Pro Arg
                    245                    250                    255
Asp His Ser Leu Asn Ala Ser Gln Gly His Phe Met Phe Ile Leu
                    260                    265                    270
Lys Lys Ser Ser Ser Leu Trp Gln Val Ala Lys Leu Gln Ser Pro
                    275                    280                    285
Thr Phe Ser Gln Thr Gly Pro Gly Cys Ile Leu Ser Phe Trp Phe
                    290                    295                    300
Tyr Asn Tyr Gly Leu Ser Val Gly Ala Ala Glu Leu Gln Leu His
                    305                    310                    315
Met Glu Asn Ser His Asp Ser Thr Val Ile Trp Arg Val Leu Tyr
                    320                    325                    330
Asn Gln Gly Lys Gln Trp Leu Glu Ala Thr Ile Gln Leu Gly Arg
                    335                    340                    345
Leu Ser Gln Pro Phe His Leu Ser Leu Asp Lys Val Ser Leu Gly
                    350                    355                    360
Ile Tyr Asp Gly Val Ser Ala Ile Asp Asp Ile Arg Phe Glu Asn
                    365                    370                    375
Cys Thr Leu Pro Leu Pro Ala Glu Ser Cys Glu Gly Leu Asp His
                    380                    385                    390
Phe Trp Cys Arg His Thr Arg Ala Cys Ile Glu Lys Leu Arg Leu
                    395                    400                    405
Cys Asp Leu Val Asp Asp Cys Gly Asp Arg Thr Asp Glu Val Asn
                    410                    415                    420
Cys Ala Pro Glu Leu Gln Cys Asn Phe Glu Thr Gly Ile Cys Asn
                    425                    430                    435
Trp Glu Gln Asp Ala Lys Asp Asp Phe Asp Trp Thr Arg Asn Gln
                    440                    445                    450
Gly Pro Thr Pro Thr Leu Asn Thr Gly Pro Met Lys Asp Asn Thr
                    455                    460                    465
Leu Gly Thr Ala Lys Gly His Tyr Leu Tyr Ile Glu Ser Ser Glu
                    470                    475                    480
Pro Gln Ala Phe Gln Asp Ser Ala Ala Leu Leu Ser Pro Ile Leu
                    485                    490                    495
Asn Ala Thr Asp Thr Lys Gly Cys Thr Phe Arg Phe Tyr Tyr His
                    500                    505                    510
Met Phe Gly Lys Arg Ile Tyr Arg Leu Ala Ile Tyr Gln Arg Ile
                    515                    520                    525
Trp Ser Asp Ser Arg Gly Gln Leu Leu Trp Gln Ile Phe Gly Asn
                    530                    535                    540
Gln Gly Asn Arg Trp Ile Arg Lys His Leu Asn Ile Ser Ser Arg
                    545                    550                    555
Gln Pro Phe Gln Ile Leu Val Glu Ala Ser Val Gly Asp Gly Phe
                    560                    565                    570
Thr Gly Asp Ile Ala Ile Asp Asp Leu Ser Phe Met Asp Cys Thr
                    575                    580                    585
Leu Tyr Pro Gly Asn Leu Pro Ala Asp Leu Pro Thr Pro Pro Glu
                    590                    595                    600
Thr Ser Val Pro Val Thr Leu Pro Pro His Asn Cys Thr Asp Ser
                    605                    610                    615
Glu Phe Ile Cys Arg Ser Asp Gly His Cys Ile Glu Lys Met Gln
                    620                    625                    630
Lys Cys Asp Phe Lys Tyr Asp Cys Pro Asp Lys Ser Asp Glu Ala
                    635                    640                    645
Ser Cys Val Met Glu Val Cys Ser Phe Glu Lys Arg Ser Leu Cys
```

```
              650                 655                 660
Lys Trp Tyr Gln Pro Ile Pro Val His Leu Leu Gln Asp Ser Asn
              665                 670                 675
Thr Phe Arg Trp Gly Leu Gly Asn Gly Ile Ser Ile His His Gly
              680                 685                 690
Glu Glu Asn His Arg Pro Ser Val Asp His Thr Gln Asn Thr Thr
              695                 700                 705
Asp Gly Trp Tyr Leu Tyr Ala Asp Ser Ser Asn Gly Lys Phe Gly
              710                 715                 720
Asp Thr Ala Asp Ile Leu Thr Pro Ile Ile Ser Leu Thr Gly Pro
              725                 730                 735
Lys Cys Thr Leu Val Phe Trp Thr His Met Asn Gly Ala Thr Val
              740                 745                 750
Gly Ser Leu Gln Val Leu Ile Lys Lys Asp Asn Val Thr Ser Lys
              755                 760                 765
Leu Trp Ala Gln Thr Gly Gln Gln Gly Ala Gln Trp Lys Arg Ala
              770                 775                 780
Glu Val Phe Leu Gly Ile Arg Ser His Thr Gln Ile Val Phe Arg
              785                 790                 795
Ala Lys Arg Gly Ile Ser Tyr Ile Gly Asp Val Ala Val Asp Asp
              800                 805                 810
Ile Ser Phe Gln Asp Cys Ser Pro Leu Leu Ser Pro Glu Arg Lys
              815                 820                 825
Cys Thr Asp His Glu Phe Met Cys Ala Asn Lys His Cys Ile Ala
              830                 835                 840
Lys Asp Lys Leu Cys Asp Phe Val Asn Asp Cys Ala Asp Asn Ser
              845                 850                 855
Asp Glu Thr Thr Phe Ile Cys Arg Thr Ser Ser Gly Arg Cys Asp
              860                 865                 870
Phe Glu Phe Asp Leu Cys Ser Trp Lys Gln Glu Lys Asp Glu Asp
              875                 880                 885
Phe Asp Trp Asn Leu Lys Ala Ser Ser Ile Pro Ala Ala Gly Thr
              890                 895                 900
Glu Pro Ala Ala Asp His Thr Leu Gly Asn Ser Ser Gly His Tyr
              905                 910                 915
Ile Phe Ile Lys Ser Leu Phe Pro Gln Gln Pro Met Arg Ala Ala
              920                 925                 930
Arg Ile Ser Ser Pro Val Ile Ser Lys Arg Ser Lys Asn Cys Lys
              935                 940                 945
Ile Ile Phe His Tyr His Met Tyr Gly Asn Gly Ile Gly Ala Leu
              950                 955                 960
Thr Leu Met Gln Val Ser Val Thr Asn Gln Thr Lys Val Leu Leu
              965                 970                 975
Asn Leu Thr Val Glu Gln Gly Asn Phe Trp Arg Arg Glu Glu Leu
              980                 985                 990
Ser Leu Phe Gly Asp Glu Asp Phe Gln Leu Lys Phe Glu Gly Arg
              995                1000                1005
Val Gly Lys Gly Gln Arg Gly Asp Ile Ala Leu Asp Asp Ile Val
             1010                1015                1020
Leu Thr Glu Asn Cys Leu Ser Leu His Asp Ser Val Gln Glu Glu
             1025                1030                1035
Leu Ala Val Pro Leu Pro Thr Gly Phe Cys Pro Leu Gly Tyr Arg
             1040                1045                1050
Glu Cys His Asn Gly Lys Cys Tyr Arg Leu Glu Gln Ser Cys Asn
             1055                1060                1065
Phe Val Asp Asn Cys Gly Asp Asn Thr Asp Glu Asn Glu Cys Gly
```

```
                    1070                    1075                    1080
Ser Ser Cys Thr Phe Glu Lys Gly Trp Cys Gly Trp Gln Asn Ser
               1085                    1090                    1095
Gln Ala Asp Asn Phe Asp Trp Val Leu Gly Val Gly Ser His Gln
               1100                    1105                    1110
Ser Leu Arg Pro Pro Lys Asp His Thr Leu Gly Asn Glu Asn Gly
               1115                    1120                    1125
His Phe Met Tyr Leu Glu Ala Thr Ala Val Gly Leu Arg Gly Asp
               1130                    1135                    1140
Lys Ala His Phe Arg Ser Thr Met Trp Arg Glu Ser Ser Ala Ala
               1145                    1150                    1155
Cys Thr Met Ser Phe Trp Tyr Phe Ile Ser Ala Lys Ala Thr Gly
               1160                    1165                    1170
Ser Ile Gln Ile Leu Ile Lys Thr Glu Lys Gly Leu Ser Lys Val
               1175                    1180                    1185
Trp Gln Glu Ser Lys Gln Asn Pro Gly Asn His Trp Gln Lys Ala
               1190                    1195                    1200
Asp Ile Leu Leu Gly Lys Leu Arg Asn Phe Glu Val Ile Phe Gln
               1205                    1210                    1215
Gly Ile Arg Thr Arg Asp Leu Gly Gly Gly Ala Ala Ile Asp Asp
               1220                    1225                    1230
Ile Glu Phe Lys Asn Cys Thr Thr Val Gly Glu Ile Ser Glu Leu
               1235                    1240                    1245
Cys Pro Glu Ile Thr Asp Phe Leu Cys Arg Asp Lys Lys Cys Ile
               1250                    1255                    1260
Ala Ser His Leu Leu Cys Asp Tyr Lys Pro Asp Cys Ser Asp Arg
               1265                    1270                    1275
Ser Asp Glu Ala His Cys Ala His Tyr Thr Ser Thr Thr Gly Ser
               1280                    1285                    1290
Cys Asn Phe Glu Thr Ser Ser Gly Asn Trp Thr Thr Ala Cys Ser
               1295                    1300                    1305
Leu Thr Gln Asp Ser Glu Asp Asp Leu Asp Trp Ala Ile Gly Ser
               1310                    1315                    1320
Arg Ile Pro Ala Lys Ala Leu Ile Pro Asp Ser Asp His Thr Pro
               1325                    1330                    1335
Gly Ser Gly Gln His Phe Leu Tyr Val Asn Ser Ser Gly Ser Lys
               1340                    1345                    1350
Glu Gly Ser Val Ala Arg Ile Thr Thr Ser Lys Ser Phe Pro Ala
               1355                    1360                    1365
Ser Leu Gly Met Cys Thr Val Arg Phe Trp Phe Tyr Met Ile Asp
               1370                    1375                    1380
Pro Arg Ser Met Gly Ile Leu Lys Val Tyr Thr Ile Glu Glu Ser
               1385                    1390                    1395
Gly Leu Asn Ile Leu Val Trp Ser Val Ile Gly Asn Lys Arg Thr
               1400                    1405                    1410
Gly Trp Thr Tyr Gly Ser Val Pro Leu Ser Ser Asn Ser Pro Phe
               1415                    1420                    1425
Lys Val Ala Phe Glu Ala Asp Leu Asp Gly Asn Glu Asp Ile Phe
               1430                    1435                    1440
Ile Ala Leu Asp Asp Ile Ser Phe Thr Pro Glu Cys Val Thr Gly
               1445                    1450                    1455
Gly Pro Val Pro Val Gln Pro Ser Pro Cys Glu Ala Asp Gln Phe
               1460                    1465                    1470
Ser Cys Ile Tyr Thr Leu Gln Cys Val Pro Leu Ser Gly Lys Cys
               1475                    1480                    1485
Asp Gly His Glu Asp Cys Ile Asp Gly Ser Asp Glu Met Asp Cys
```

```
                 1490                1495                1500
Pro Leu Ser Pro Thr Pro Pro Leu Cys Ser Asn Met Glu Phe Pro
             1505                1510                1515
Cys Ser Thr Asp Glu Cys Ile Pro Ser Leu Leu Leu Cys Asp Gly
             1520                1525                1530
Val Pro Asp Cys His Phe Asn Glu Asp Glu Leu Ile Cys Ser Asn
             1535                1540                1545
Lys Ser Cys Ser Asn Gly Ala Leu Val Cys Ala Ser Ser Asn Ser
             1550                1555                1560
Cys Ile Pro Ala His Gln Arg Cys Asp Gly Phe Ala Asp Cys Met
             1565                1570                1575
Asp Phe Gln Leu Asp Glu Ser Ser Cys Ser Glu Cys Pro Leu Asn
             1580                1585                1590
Tyr Cys Arg Asn Gly Gly Thr Cys Val Val Glu Lys Asn Gly Pro
             1595                1600                1605
Met Cys Arg Cys Arg Gln Gly Trp Lys Gly Asn Arg Cys His Ile
             1610                1615                1620
Lys Phe Asn Pro Pro Ala Thr Asp Phe Thr Tyr Ala Gln Asn Asn
             1625                1630                1635
Thr Trp Thr Leu Leu Gly Ile Gly Leu Ala Phe Leu Met Thr His
             1640                1645                1650
Ile Thr Val Ala Val Leu Cys Phe Leu Ala Asn Arg Lys Val Pro
             1655                1660                1665
Ile Arg Lys Thr Glu Gly Ser Gly Asn Cys Ala Phe Val Asn Pro
             1670                1675                1680
Val Tyr Gly Asn Trp Ser Asn Pro Glu Lys Thr Glu Ser Ser Val
             1685                1690                1695
Tyr Ser Phe Ser Asn Pro Leu Tyr Gly Thr Thr Ser Gly Ser Leu
             1700                1705                1710
Glu Thr Leu Ser His His Leu Lys
             1715
```

```
<210> 7
<211> 224
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 382654CD1

<400> 7
Met Leu Leu Ser Pro Asp Gln Lys Val Leu Thr Ile Thr Arg Val
 1               5                10                15
Leu Met Glu Asp Asp Asp Leu Tyr Ser Cys Met Val Glu Asn Pro
             20                25                30
Ile Ser Gln Gly Arg Ser Leu Pro Val Lys Ile Thr Val Tyr Arg
             35                40                45
Arg Ser Ser Leu Tyr Ile Ile Leu Ser Thr Gly Gly Ile Phe Leu
             50                55                60
Leu Val Thr Leu Val Thr Val Cys Ala Cys Trp Lys Pro Ser Lys
             65                70                75
Arg Lys Gln Lys Lys Leu Glu Lys Gln Asn Ser Leu Glu Tyr Met
             80                85                90
Asp Gln Asn Asp Asp Arg Leu Lys Pro Glu Ala Asp Thr Leu Pro
             95                100               105
```

```
Arg Ser Gly Glu Gln Glu Arg Lys Asn Pro Met Ala Leu Tyr Ile
            110                 115                 120
Leu Lys Asp Lys Asp Ser Pro Glu Thr Glu Glu Asn Pro Ala Pro
            125                 130                 135
Glu Pro Arg Ser Ala Thr Glu Pro Gly Pro Pro Gly Tyr Ser Val
             140                 145                 150
Ser Pro Ala Val Pro Gly Arg Ser Pro Gly Leu Pro Ile Arg Ser
            155                 160                 165
Ala Arg Arg Tyr Pro Arg Ser Pro Ala Arg Ser Pro Ala Thr Gly
            170                 175                 180
Arg Thr His Ser Ser Pro Pro Arg Ala Pro Ser Ser Pro Gly Arg
            185                 190                 195
Ser Arg Ser Ala Ser Arg Thr Leu Arg Thr Ala Gly Val His Ile
            200                 205                 210
Ile Arg Glu Gln Asp Glu Ala Gly Pro Val Glu Ile Ser Ala
            215                 220
```

```
<210> 8
<211> 570
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1867351CD1

<400> 8
Met Glu Ala Pro Glu Glu Pro Ala Pro Val Arg Gly Gly Pro Glu
 1               5                  10                  15
Ala Thr Leu Glu Val Arg Gly Ser Arg Cys Leu Arg Leu Ser Ala
            20                  25                  30
Phe Arg Glu Glu Leu Arg Ala Leu Leu Val Leu Ala Gly Pro Ala
            35                  40                  45
Phe Leu Val Gln Leu Met Val Phe Leu Ile Ser Phe Ile Ser Ser
            50                  55                  60
Val Phe Cys Gly His Leu Gly Lys Leu Glu Leu Asp Ala Val Thr
            65                  70                  75
Leu Ala Ile Ala Val Ile Asn Val Thr Gly Val Ser Val Gly Phe
            80                  85                  90
Gly Leu Ser Ser Ala Cys Asp Thr Leu Ile Ser Gln Thr Tyr Gly
            95                 100                 105
Ser Gln Asn Leu Lys His Val Gly Val Ile Leu Gln Arg Ser Ala
            110                 115                 120
Leu Val Leu Leu Leu Cys Cys Phe Pro Cys Trp Ala Leu Phe Leu
            125                 130                 135
Asn Thr Gln His Ile Leu Leu Leu Phe Arg Gln Asp Pro Asp Val
            140                 145                 150
Ser Arg Leu Thr Gln Thr Tyr Val Thr Ile Phe Ile Pro Ala Leu
            155                 160                 165
Pro Ala Thr Phe Leu Tyr Met Leu Gln Val Lys Tyr Leu Leu Asn
            170                 175                 180
Gln Gly Ile Val Leu Pro Gln Ile Val Thr Gly Val Ala Ala Asn
            185                 190                 195
Leu Val Asn Ala Leu Ala Asn Tyr Leu Phe Leu His Gln Leu His
            200                 205                 210
Leu Gly Val Ile Gly Ser Ala Leu Ala Asn Leu Ile Ser Gln Tyr
```

```
                215                       220                       225
    Thr Leu Ala Leu Leu Leu Phe Leu Tyr Ile Leu Gly Lys Lys Leu
                230                       235                       240
    His Gln Ala Thr Trp Gly Gly Trp Ser Leu Glu Cys Leu Gln Asp
                245                       250                       255
    Trp Ala Ser Phe Leu Arg Leu Ala Ile Pro Ser Met Leu Met Leu
                260                       265                       270
    Cys Met Glu Trp Trp Ala Tyr Glu Val Gly Ser Phe Pro Ser Gly
                275                       280                       285
    Ile Leu Gly Met Val Glu Leu Gly Ala Gln Ser Ile Val Tyr Glu
                290                       295                       300
    Leu Ala Ile Ile Val Tyr Met Val Pro Ala Asp Phe Ser Val Ala
                305                       310                       315
    Ala Ser Val Arg Val Gly Asn Ala Leu Gly Ala Gly Asp Met Glu
                320                       325                       330
    Gln Ala Arg Lys Ser Ser Thr Val Ser Leu Leu Ile Thr Val Leu
                335                       340                       345
    Phe Ala Val Ala Phe Ser Val Leu Leu Leu Ser Cys Lys Asp His
                350                       355                       360
    Val Gly Tyr Ile Phe Thr Thr Asp Arg Asp Ile Ile Asn Leu Val
                365                       370                       375
    Ala Gln Val Val Pro Ile Tyr Ala Val Ser His Leu Phe Glu Ala
                380                       385                       390
    Leu Ala Cys Thr Ser Gly Gly Val Leu Arg Gly Ser Gly Asn Gln
                395                       400                       405
    Lys Val Gly Ala Ile Val Asn Thr Ile Gly Tyr Tyr Val Val Gly
                410                       415                       420
    Leu Pro Ile Gly Ile Ala Leu Met Phe Ala Thr Thr Leu Gly Val
                425                       430                       435
    Met Gly Leu Trp Ser Gly Ile Ile Ile Cys Thr Val Phe Gln Ala
                440                       445                       450
    Val Cys Phe Leu Gly Phe Ile Ile Gln Leu Asn Trp Lys Lys Ala
                455                       460                       465
    Cys Gln Gln Ala Gln Val His Ala Asn Leu Lys Val Asn Asn Val
                470                       475                       480
    Pro Arg Ser Gly Asn Ser Ala Leu Pro Gln Asp Pro Leu His Pro
                485                       490                       495
    Gly Cys Pro Glu Asn Leu Glu Gly Ile Leu Thr Asn Asp Val Gly
                500                       505                       510
    Lys Thr Gly Glu Pro Gln Ser Asp Gln Gln Met Arg Gln Glu Glu
                515                       520                       525
    Pro Leu Pro Glu His Pro Gln Asp Gly Ala Lys Leu Ser Arg Lys
                530                       535                       540
    Gln Leu Val Leu Arg Arg Gly Leu Leu Leu Leu Gly Val Phe Leu
                545                       550                       555
    Ile Leu Leu Val Gly Ile Leu Val Arg Phe Tyr Val Arg Ile Gln
                560                       565                       570
```

<210> 9
<211> 423
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 3323104CD1

<400> 9

Met	Gly	Ser	Thr	Lys	His	Trp	Gly	Glu	Leu	Leu	Leu	Asn	Leu	Lys
1				5					10					15

Val Ala Pro Ala Gly Val Phe Gly Val Ala Phe Leu Ala Arg Val
                  20   25   30

Ala Leu Val Phe Tyr Gly Val Phe Gln Asp Arg Thr Leu His Val
                  35   40   45

Arg Tyr Thr Asp Ile Asp Tyr Gln Val Phe Thr Asp Ala Ala Arg
                  50   55   60

Phe Val Thr Glu Gly Arg Ser Pro Tyr Leu Arg Ala Thr Tyr Arg
                  65   70   75

Tyr Thr Pro Leu Leu Gly Trp Leu Leu Thr Pro Asn Ile Tyr Leu
                  80   85   90

Ser Glu Leu Phe Gly Lys Phe Leu Phe Ile Ser Cys Asp Leu Leu
                  95   100   105

Thr Ala Phe Leu Leu Tyr Arg Leu Leu Leu Leu Lys Gly Leu Gly
                  110   115   120

Arg Arg Gln Ala Cys Gly Tyr Cys Val Phe Trp Leu Leu Asn Pro
                  125   130   135

Leu Pro Met Ala Val Ser Ser Arg Gly Asn Ala Asp Ser Ile Val
                  140   145   150

Ala Ser Leu Val Leu Met Val Leu Tyr Leu Ile Lys Lys Arg Leu
                  155   160   165

Val Ala Cys Ala Ala Val Phe Tyr Gly Phe Ala Val His Met Lys
                  170   175   180

Ile Tyr Pro Val Thr Tyr Ile Leu Pro Ile Thr Leu His Leu Leu
                  185   190   195

Pro Asp Arg Asp Asn Asp Lys Ser Leu Arg Gln Phe Arg Tyr Thr
                  200   205   210

Phe Gln Ala Cys Leu Tyr Glu Leu Leu Lys Arg Leu Cys Asn Arg
                  215   220   225

Ala Val Leu Leu Phe Val Ala Val Ala Gly Leu Thr Phe Phe Ala
                  230   235   240

Leu Ser Phe Gly Phe Tyr Tyr Glu Tyr Gly Trp Glu Phe Leu Glu
                  245   250   255

His Thr Tyr Phe Tyr His Leu Thr Arg Arg Asp Ile Arg His Asn
                  260   265   270

Phe Ser Pro Tyr Phe Tyr Met Leu Tyr Leu Thr Ala Glu Ser Lys
                  275   280   285

Trp Ser Phe Ser Leu Gly Ile Ala Ala Phe Leu Pro Gln Leu Ile
                  290   295   300

Leu Leu Ser Ala Val Ser Phe Ala Tyr Tyr Arg Asp Leu Val Phe
                  305   310   315

Cys Cys Phe Leu His Thr Ser Ile Phe Val Thr Phe Asn Lys Val
                  320   325   330

Cys Thr Ser Gln Tyr Phe Leu Trp Tyr Leu Cys Leu Leu Pro Leu
                  335   340   345

Val Met Pro Leu Val Arg Met Pro Trp Lys Arg Ala Val Val Leu
                  350   355   360

Leu Met Leu Trp Leu Ile Gly Gln Ala Met Trp Leu Ala Pro Ala
                  365   370   375

Tyr Val Leu Glu Phe Gln Gly Lys Asn Thr Phe Leu Phe Ile Trp
                  380   385   390

Leu Ala Gly Leu Phe Phe Leu Leu Ile Asn Cys Ser Ile Leu Ile

```
                    395                    400                    405
Gln Ile Ile Ser His Tyr Lys Glu Glu Pro Leu Thr Glu Arg Ile
                    410                    415                    420
Lys Tyr Asp
```

<210> 10
<211> 388
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4769306CD1

<400> 10
```
Met Gly Phe Ser Ala Arg Tyr Asn Phe Thr Pro Asp Pro Asp Phe
  1           5                  10                     15
Lys Asp Leu Gly Ala Leu Lys Pro Leu Pro Ala Cys Glu Phe Glu
                 20                  25                     30
Met Gly Gly Ser Glu Gly Ile Val Glu Ser Ile Gln Ile Met Lys
                 35                  40                     45
Glu Gly Lys Ala Thr Ala Ser Glu Ala Val Asp Cys Lys Trp Tyr
                 50                  55                     60
Ile Arg Ala Pro Pro Arg Ser Lys Ile Tyr Leu Arg Phe Leu Asp
                 65                  70                     75
Tyr Glu Met Gln Asn Ser Asn Glu Cys Lys Arg Asn Phe Val Ala
                 80                  85                     90
Val Tyr Asp Gly Ser Ser Ser Val Glu Asp Leu Lys Ala Lys Phe
                 95                 100                    105
Cys Ser Thr Val Ala Asn Asp Val Met Leu Arg Thr Gly Leu Gly
                110                 115                    120
Val Ile Arg Met Trp Ala Asp Glu Gly Ser Arg Asn Ser Arg Phe
                125                 130                    135
Gln Met Leu Phe Thr Ser Phe Gln Glu Pro Pro Cys Glu Gly Asn
                140                 145                    150
Thr Phe Phe Cys His Ser Asn Met Cys Ile Asn Asn Thr Leu Val
                155                 160                    165
Cys Asn Gly Leu Gln Asn Cys Val Tyr Pro Trp Asp Glu Asn His
                170                 175                    180
Cys Lys Glu Lys Arg Lys Thr Ser Leu Leu Asp Gln Leu Thr Asn
                185                 190                    195
Thr Ser Gly Thr Val Ile Gly Val Thr Ser Cys Ile Val Ile Ile
                200                 205                    210
Leu Ile Ile Ile Ser Val Ile Val Gln Ile Lys Gln Pro Arg Lys
                215                 220                    225
Lys Tyr Val Gln Arg Lys Ser Asp Phe Asp Gln Thr Val Phe Gln
                230                 235                    240
Glu Val Phe Glu Pro Pro His Tyr Glu Leu Cys Thr Leu Arg Gly
                245                 250                    255
Thr Gly Ala Thr Ala Asp Phe Ala Asp Val Ala Asp Asp Phe Glu
                260                 265                    270
Asn Tyr His Lys Leu Arg Arg Ser Ser Ser Lys Cys Ile His Asp
                275                 280                    285
His His Cys Gly Ser Gln Leu Ser Ser Thr Lys Gly Ser Arg Ser
                290                 295                    300
```

Asn Leu Ser Thr Arg Asp Ala Ser Ile Leu Thr Glu Met Pro Thr
              305   310   315
Gln Pro Gly Lys Pro Leu Ile Pro Pro Met Asn Arg Arg Asn Ile
              320   325   330
Leu Val Met Lys His Asn Tyr Ser Gln Asp Ala Ala Asp Ala Cys
              335   340   345
Asp Ile Asp Glu Ile Glu Glu Val Pro Thr Thr Ser His Arg Leu
              350   355   360
Ser Arg His Asp Lys Ala Val Gln Arg Phe Cys Leu Ile Gly Ser
              365   370   375
Leu Ser Lys His Glu Ser Glu Tyr Asn Thr Thr Arg Val
              380   385

<210> 11
<211> 231
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2720058CD1

<400> 11
Met Ala Phe Val Pro Phe Leu Leu Val Thr Trp Ser Ser Ala Ala
1   5   10   15
Phe Ile Ile Ser Tyr Val Val Ala Val Leu Ser Gly His Val Asn
              20   ·25   30
Pro Phe Leu Pro Tyr Ile Ser Asp Thr Gly Thr Thr Pro Pro Glu
              35   40   ·   · 45
Ser Gly Ile Phe Gly Phe Met Ile Asn ·Phe Ser Ala Phe Leu Gly
       ·   50   55   ·   60
Ala Ala Thr Met Tyr Thr Arg Tyr Lys Ile Val Gln Lys Gln Asn
              65   · 70   75
Gln Thr Cys Tyr Phe Ser Thr Pro Val Phe Asn Leu Val· Ser Leu
              80   . 85   90
Val Leu Gly Leu Val Gly Cys Phe Gly Met Gly Ile Val Ala Asn
              95   100   105
Phe Gln Glu Leu Ala Val Pro Val Val His Asp Gly Gly Ala Leu
              110   115   120
Leu Ala Phe Val·Cys Gly Val Val Tyr Thr Leu Leu Gln Ser Ile
              125   130   135
Ile Ser Tyr Lys Ser Cys Pro Gln Trp Asn Ser Leu Ser Thr Cys
              140   145   150
His Ile Arg Met Val Ile Ser Ala Val Ser Cys Ala Ala Val Ile
              155   160   165
Pro Met Ile Val Cys Ala Ser Leu Ile Ser Ile Thr Lys Leu Glu
              170   175   180
Trp Asn Pro Arg Glu Lys Asp Tyr Val Tyr His Val Val Ser Ala
              185   190   195
Ile Cys Glu Trp Thr Val Ala Phe Gly Phe Ile Phe Tyr Phe Leu
              200   205   210
Thr Phe Ile Gln Asp Phe Gln Ser Val Thr Leu Arg Ile Ser Thr
              215   220   225
Glu Ile Asn Gly Asp Ile
              230

```
<210> 12
<211> 293
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481255CD1

<400> 12
Met Asp Arg Ala Lys Gln Gln Gln Ala Leu Leu Leu Leu Pro Val
  1               5                  10                  15
Cys Leu Ala Leu Thr Phe Ser Leu Thr Ala Val Val Ser Ser His
               20                  25                  30
Trp Cys Glu Gly Thr Arg Arg Val Val Lys Pro Leu Cys Gln Asp
               35                  40                  45
Gln Pro Gly Gly Gln His Cys Ile His Phe Lys Arg Asp Asn Ser
               50                  55                  60
Ser Asn Gly Arg Met Asp Asn Asn Ser Gln Ala Val Leu Tyr Ile
               65                  70                  75
Trp Glu Leu Gly Asp Asp Lys Phe Ile Gln Arg Gly Phe His Val
               80                  85                  90
Gly Leu Trp Gln Ser Cys Glu Glu Ser Leu Asn Gly Glu Asp Glu
               95                 100                 105
Lys Cys Arg Ser Phe Arg Ser Val Val Pro Ala Glu Glu Gln Gly
              110                 115                 120
Val Leu Trp Leu Ser Ile Gly Gly Glu Val Leu Asp Ile Val Leu
              125                 130                 135
Ile Leu Thr Ser Ala Ile Leu Leu Gly Ser Arg Val Ser Cys Arg
              140                 145                 150
Ser Pro Gly Phe His Trp Leu Arg Val Asp Ala Leu Val Ala Ile
              155                 160                 165
Phe Met Val Leu Ala Gly Leu Leu Gly Met Val Ala His Met Met
              170                 175                 180
Tyr Thr Thr Ile Phe Gln Ile Thr Val Asn Leu Gly Pro Glu Asp
              185                 190                 195
Trp Lys Pro Gln Thr Trp Asp Tyr Gly Trp Ser Tyr Cys Leu Ala
              200                 205                 210
Trp Gly Ser Phe Ala Leu Cys Leu Ala Val Ser Val Ser Ala Met
              215                 220                 225
Ser Arg Phe Thr Ala Ala Arg Leu Glu Phe Thr Glu Lys Gln Gln
              230                 235                 240
Ala Gln Asn Gly Ser Arg His Ser Gln His Ser Phe Leu Glu Pro
              245                 250                 255
Glu Ala Ser Glu Ser Ile Trp Lys Thr Gly Ala Ala Pro Cys Pro
              260                 265                 270
Ala Glu Gln Ala Phe Arg Asn Val Ser Gly His Leu Pro Pro Gly
              275                 280                 285
Ala Pro Gly Lys Val Ser Ile Cys
              290

<210> 13
<211> 526
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> misc_feature
<223> Incyte ID No: 1510242CD1


<400> 13
Met Leu Thr Tyr Gly Val Tyr Leu Gly Leu Leu Gln Met Gln Leu
 1               5                   10                  15
Ile Leu His Tyr Asp Glu Thr Tyr Arg Glu Val Lys Tyr Gly Asn
                20                  25                  30
Met Gly Leu Pro Asp Ile Asp Ser Lys Met Leu Met Gly Ile Asn
                35                  40                  45
Val Thr Pro Ile Ala Ala Leu Leu Tyr Thr Pro Val Leu Ile Arg
                50                  55                  60
Phe Phe Gly Thr Lys Trp Met Met Phe Leu Ala Val Gly Ile Tyr
                65                  70                  75
Ala Leu Phe Val Ser Thr Asn Tyr Trp Glu Arg Tyr Tyr Thr Leu
                80                  85                  90
Val Pro Ser Ala Val Ala Leu Gly Met Ala Ile Val Pro Leu Trp
                95                  100                 105
Ala Ser Met Gly Asn Tyr Ile Thr Arg Met Ala Gln Lys Tyr His
                110                 115                 120
Glu Tyr Ser His Tyr Lys Glu Gln Asp Gly Gln Gly Met Lys Gln
                125                 130                 135
Arg Pro Pro Arg Gly Ser His Ala Pro Tyr Leu Leu Val Phe Gln
                140                 145                 150
Ala Ile Phe Tyr Ser Phe Phe His Leu Ser Phe Ala Cys Ala Gln
                155                 160                 165
Leu Pro Met Ile Tyr Phe Leu Asn His Tyr Leu Tyr Asp Leu Asn
                170                 175                 180
His Thr Leu Tyr Asn Val Gln Ser Cys Gly Thr Asn Ser His Gly
                185                 190                 195
Ile Leu Ser Gly Phe Asn Lys Thr Val Leu Arg Thr Leu Pro Arg
                200                 205                 210
Ser Gly Asn Leu Ile Val Val Glu Ser Val Leu Met Ala Val Ala
                215                 220                 225
Phe Leu Ala Met Leu Leu Val Leu Gly Leu Cys Gly Ala Ala Tyr
                230                 235                 240
Arg Pro Thr Glu Glu Ile Asp Leu Arg Ser Val Gly Trp Gly Asn
                245                 250                 255
Ile Phe Gln Leu Pro Phe Lys His Val Arg Asp Tyr Arg Leu Arg
                260                 265                 270
His Leu Val Pro Phe Phe Ile Tyr Ser Gly Phe Glu Val Leu Phe
                275                 280                 285
Ala Cys Thr Gly Ile Ala Leu Gly Tyr Gly Val Cys Ser Val Gly
                290                 295                 300
Leu Glu Arg Leu Ala Tyr Leu Leu Val Ala Tyr Ser Leu Gly Ala
                305                 310                 315
Ser Ala Ala Ser Leu Leu Gly Leu Leu Gly Leu Trp Leu Pro Arg
                320                 325                 330
Pro Val Pro Leu Val Ala Gly Ala Gly Val His Leu Leu Leu Thr
                335                 340                 345
Phe Ile Leu Phe Phe Trp Ala Pro Val Pro Arg Val Leu Gln His
                350                 355                 360
Ser Trp Ile Leu Tyr Val Ala Ala Ala Leu Trp Gly Val Gly Ser
                365                 370                 375
Ala Leu Asn Lys Thr Gly Leu Ser Thr Leu Leu Gly Ile Leu Tyr
```

```
                    380                385                390
Glu Asp Lys Glu Arg Gln Asp Phe Ile Phe Thr Ile Tyr His Trp
                    395                400                405
Trp Gln Ala Val Ala Ile Phe Thr Val Tyr Leu Gly Ser Ser Leu
                    410                415                420
His Met Lys Ala Lys Leu Ala Val Leu Leu Val Thr Leu Val Ala
                    425                430                435
Ala Ala Val Ser Tyr Leu Arg Met Glu Gln Lys Leu Arg Arg Gly
                    440                445                450
Val Ala Pro Arg Gln Pro Arg Ile Pro Arg Pro Gln His Lys Val
                    455                460                465
Arg Gly Tyr Arg Tyr Leu Glu Glu Asp Asn Ser Asp Glu Ser Asp
                    470                475                480
Ala Glu Gly Glu His Gly Asp Gly Ala Glu Glu Glu Ala Pro Pro
                    485                490                495
Ala Gly Pro Arg Pro Gly Pro Glu Pro Ala Gly Leu Gly Arg Arg
                    500                505                510
Pro Cys Pro Tyr Glu Gln Ala Gln Gly Gly Asp Gly Pro Glu Glu
                    515                520                525
Gln
```

```
<210> 14
<211> 348
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 162131CD1

<400> 14
Met Gly Ser Trp Val Gln Leu Ile Thr Ser Val Gly Val Gln Gln
 1               5                10                15
Asn His Pro Gly Trp Thr Val Ala Gly Gln Phe Gln Glu Lys Lys
                    20                25                30
Arg Phe Thr Glu Glu Val Ile Glu Tyr Phe Gln Lys Lys Val Ser
                    35                40                45
Pro Val His Leu Lys Ile Leu Leu Thr Ser Asp Glu Ala Trp Lys
                    50                55                60
Arg Phe Val Arg Val Ala Glu Leu Pro Arg Glu Glu Ala Asp Ala
                    65                70                75
Leu Tyr Glu Ala Leu Lys Asn Leu Thr Pro Tyr Val Ala Ile Glu
                    80                85                90
Asp Lys Asp Met Gln Gln Lys Glu Gln Gln Phe Arg Glu Trp Phe
                    95                100               105
Leu Lys Glu Phe Pro Gln Ile Arg Trp Lys Ile Gln Glu Ser Ile
                    110               115               120
Glu Arg Leu Arg Val Ile Ala Asn Glu Ile Glu Lys Val His Arg
                    125               130               135
Gly Cys Val Ile Ala Asn Val Val Ser Gly Ser Thr Gly Ile Leu
                    140               145               150
Ser Val Ile Gly Val Met Leu Ala Pro Phe Thr Ala Gly Leu Ser
                    155               160               165
Leu Ser Ile Thr Ala Ala Gly Val Gly Leu Gly Ile Ala Ser Ala
                    170               175               180
```

```
Thr Ala Gly Ile Ala Ser Ser Ile Val Glu Asn Thr Tyr Thr Arg
            185                 190                 195
Ser Ala Glu Leu Thr Ala Ser Arg Leu Thr Ala Thr Ser Thr Asp
            200                 205                 210
Gln Leu Glu Ala Leu Arg Asp Ile Leu His Asp Ile Thr Pro Asn
            215                 220                 225
Val Leu Ser Phe Ala Leu Asp Phe Asp Glu Ala Thr Lys Met Ile
            230                 235                 240
Ala Asn Asp Val His Thr Leu Arg Arg Ser Lys Ala Thr Val Gly
            245                 250                 255
Arg Pro Leu Ile Ala Trp Arg Tyr Val Pro Ile Asn Val Val Glu
            260                 265                 270
Thr Leu Arg Thr Arg Gly Ala Pro Thr Arg Ile Val Arg Lys Val
            275                 280                 285
Ala Arg Asn Leu Gly Lys Ala Thr Ser Gly Val Leu Val Val Leu
            290                 295                 300
Asp Val Val Asn Leu Val Gln Asp Ser Leu Asp Leu His Lys Gly
            305                 310                 315
Glu Lys Ser Glu Ser Ala Glu Leu Leu Arg Gln Trp Ala Gln Glu
            320                 325                 330
Leu Glu Glu Asn Leu Asn Glu Leu Thr His Ile His Gln Ser Leu
            335                 340                 345
Lys Ala Gly
```

```
<210> 15
<211> 520
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1837725CD1

<400> 15
Met Gly Pro Gln Arg Arg Leu Ser Pro Ala Gly Ala Ala Leu Leu
  1               5                  10                  15
Trp Gly Phe Leu Leu Gln Leu Thr Ala Ala Gln Glu Ala Ile Leu
            20                  25                  30
His Ala Ser Gly Asn Gly Thr Thr Lys Asp Tyr Cys Met Leu Tyr
            35                  40                  45
Asn Pro Tyr Trp Thr Ala Leu Pro Ser Thr Leu Glu Asn Ala Thr
            50                  55                  60
Ser Ile Ser Leu Met Asn Leu Thr Ser Thr Pro Leu Cys Asn Leu
            65                  70                  75
Ser Asp Ile Pro Pro Val Gly Ile Lys Ser Lys Ala Val Val Val
            80                  85                  90
Pro Trp Gly Ser Cys His Phe Leu Glu Lys Ala Arg Ile Ala Gln
            95                  100                 105
Lys Gly Gly Ala Glu Ala Met Leu Val Val Asn Asn Ser Val Leu
            110                 115                 120
Phe Pro Pro Ser Gly Asn Arg Ser Glu Phe Pro Asp Val Lys Ile
            125                 130                 135
Leu Ile Ala Phe Ile Ser Tyr Lys Asp Phe Arg Asp Met Asn Gln
            140                 145                 150
Thr Leu Gly Asp Asn Ile Thr Val Lys Met Tyr Ser Pro Ser Trp
```

```
                   155                   160                   165
Pro Asn Phe Asp Tyr Thr Met Val Val Ile Phe Val Ile Ala Val
                   170                   175                   180
Phe Thr Val Ala Leu Gly Gly Tyr Trp Ser Gly Leu Val Glu Leu
                   185                   190                   195
Glu Asn Leu Lys Ala Val Thr Thr Glu Asp Arg Glu Met Arg Lys
                   200                   205                   210
Lys Lys Glu Glu Tyr Leu Thr Phe Ser Pro Leu Thr Val Val Ile
                   215                   220                   225
Phe Val Val Ile Cys Cys Val Met Met Val Leu Leu Tyr Phe Phe
                   230                   235                   240
Tyr Lys Trp Leu Val Tyr Val Met Ile Ala Ile Phe Cys Ile Ala
                   245                   250                   255
Ser Ala Met Ser Leu Tyr Asn Cys Leu Ala Ala Leu Ile His Lys
                   260                   265                   270
Ile Pro Tyr Gly Gln Cys Thr Ile Ala Cys Arg Gly Lys Asn Met
                   275                   280                   285
Glu Val Arg Leu Ile Phe Leu Ser Gly Leu Cys Ile Ala Val Ala
                   290                   295                   300
Val Val Trp Ala Val Phe Arg Asn Glu Asp Arg Trp Ala Trp Ile
                   305                   310                   315
Leu Gln Asp Ile Leu Gly Ile Ala Phe Cys Leu Asn Leu Ile Lys
                   320                   325                   330
Thr Leu Lys Leu Pro Asn Phe Lys Ser Cys Val Ile Leu Leu Gly
                   335                   340                   345
Leu Leu Leu Leu Tyr Asp Val Phe Phe Val Phe Ile Thr Pro Phe
                   350                   355                   360
Ile Thr Lys Asn Gly Glu Ser Ile Met Val Glu Leu Ala Ala Gly
                   365                   370                   375
Pro Phe Gly Asn Asn Glu Lys Leu Pro Val Val Ile Arg Val Pro
                   380                   385                   390
Lys Leu Ile Tyr Phe Ser Val Met Ser Val Cys Leu Met Pro Val
                   395                   400                   405
Ser Ile Leu Gly Phe Gly Asp Ile Ile Val Pro Gly Leu Leu Ile
                   410                   415                   420
Ala Tyr Cys Arg Arg Phe Asp Val Gln Thr Gly Ser Ser Tyr Ile
                   425                   430                   435
Tyr Tyr Val Ser Ser Thr Val Ala Tyr Ala Ile Gly Met Ile Leu
                   440                   445                   450
Thr Phe Val Val Leu Val Leu Met Lys Lys Gly Gln Pro Ala Leu
                   455                   460                   465
Leu Tyr Leu Val Pro Cys Thr Leu Ile Thr Ala Ser Val Val Ala
                   470                   475                   480
Trp Arg Arg Lys Glu Met Lys Lys Phe Trp Lys Gly Asn Ser Tyr
                   485                   490                   495
Gln Met Met Asp His Leu Asp Cys Ala Thr Asn Glu Glu Asn Pro
                   500                   505                   510
Val Ile Ser Gly Glu Gln Ile Val Gln Gln
                   515                   520
```

```
<210> 16
<211> 534
<212> PRT
<213> Homo sapiens

<220>
```

```
<221> misc_feature
<223> Incyte ID No: 3643847CD1

<400> 16
Met Gln Ala Ala Arg Val Asp Tyr Ile Ala Pro Trp Trp Val Val
 1               5                  10                  15
Trp Leu His Ser Val Pro His Val Gly Leu Arg Leu Gln Pro Val
                20                  25                  30
Asn Ser Thr Phe Ser Pro Gly Asp Glu Ser Tyr Gln Glu Ser Leu
                35                  40                  45
Leu Phe Leu Gly Leu Val Ala Ala Val Cys Leu Gly Leu Asn Leu
                50                  55                  60
Ile Phe Leu Val Ala Tyr Leu Val Cys Ala Cys His Cys Arg Arg
                65                  70                  75
Asp Asp Ala Val Gln Thr Lys Gln His His Ser Cys Cys Ile Thr
                80                  85                  90
Trp Thr Ala Val Val Ala Gly Leu Ile Cys Cys Ala Ala Val Gly
                95                  100                 105
Val Gly Phe Tyr Gly Asn Ser Glu Thr Asn Asp Gly Ala Tyr Gln
                110                 115                 120
Leu Met Tyr Ser Leu Asp Asp Ala Asn His Thr Phe Ser Gly Ile
                125                 130                 135
Asp Ala Leu Val Ser Gly Thr Thr Gln Lys Met Lys Val Asp Leu
                140                 145                 150
Glu Gln His Leu Ala Arg Leu Ser Glu Ile Phe Ala Ala Arg Gly
                155                 160                 165
Asp Tyr Leu Gln Thr Leu Lys Phe Ile Gln Gln Met Ala Gly Ser
                170                 175                 180
Val Val Val Gln Leu Ser Gly Leu Pro Val Trp Arg Glu Val Thr
                185                 190                 195
Met Glu Leu Thr Lys Leu Ser Asp Gln Thr Gly Tyr Val Glu Tyr
                200                 205                 210
Tyr Arg Trp Leu Ser Tyr Leu Leu Leu Phe Ile Leu Asp Leu Val
                215                 220                 225
Ile Cys Leu Ile Ala Cys Leu Gly Leu Ala Lys Arg Ser Lys Cys
                230                 235                 240
Leu Leu Ala Ser Met Leu Cys Cys Gly Ala Leu Ser Leu Leu Leu
                245                 250                 255
Ser Trp Ala Ser Leu Ala Ala Asp Gly Ser Ala Ala Val Ala Thr
                260                 265                 270
Ser Asp Phe Cys Val Ala Pro Asp Thr Phe Ile Leu Asn Val Thr
                275                 280                 285
Glu Gly Gln Ile Ser Thr Glu Val Thr Arg Tyr Tyr Leu Tyr Cys
                290                 295                 300
Ser Gln Ser Gly Ser Ser Pro Phe Gln Gln Thr Leu Thr Thr Phe
                305                 310                 315
Gln Arg Ala Leu Thr Thr Met Gln Ile Gln Val Ala Gly Leu Leu
                320                 325                 330
Gln Phe Ala Val Pro Leu Phe Ser Thr Ala Glu Glu Asp Leu Leu
                335                 340                 345
Ala Ile Gln Leu Leu Leu Asn Ser Ser Glu Ser Ser Leu His Gln
                350                 355                 360
Leu Thr Ala Met Val Asp Cys Arg Gly Leu His Lys Asp Tyr Leu
                365                 370                 375
Asp Ala Leu Ala Gly Ile Cys Tyr Asp Gly Leu Gln Gly Leu Leu
                380                 385                 390
```

```
Tyr Leu Gly Leu Phe Ser Phe Leu Ala Ala Leu Ala Phe Ser Thr
              395               400               405
Met Ile Cys Ala Gly Pro Arg Ala Trp Lys His Phe Thr Thr Arg
              410               415               420
Asn Arg Glu Tyr Asp Asp Ile Asp Asp Asp Pro Phe Asn Pro
              425               430               435
Gln Ala Trp Arg Met Ala Ala His Ser Pro Pro Arg Gly Gln Leu
              440               445               450
His Ser Phe Cys Ser Tyr Ser Ser Gly Leu Gly Ser Gln Thr Ser
              455               460               465
Leu Gln Pro Pro Ala Gln Thr Ile Ser Asn Ala Pro Val Ser Glu
              470               475               480
Tyr Met Asn Gln Ala Met Leu Phe Gly Arg Asn Pro Arg Tyr Glu
              485               490               495
Asn Val Pro Leu Ile Gly Arg Ala Ser Pro Pro Pro Thr Tyr Ser
              500               505               510
Pro Ser Met Arg Ala Thr Tyr Leu Ser Val Ala Asp Glu His Leu
              515               520               525
Arg His Tyr Gly Asn Gln Phe Pro Ala
              530
```

```
<210> 17
<211> 820
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6889872CD1

<400> 17
Met Leu Arg Leu Gly Leu Cys Ala Ala Ala Leu Leu Cys Val Cys
1               5                 10                15
Arg Pro Gly Ala Val Arg Ala Asp Cys Trp Leu Ile Glu Gly Asp
              20                25                30
Lys Gly Tyr Val Trp Leu Ala Ile Cys Ser Gln Asn Gln Pro Pro
              35                40                45
Tyr Glu Thr Ile Pro Gln His Ile Asn Ser Thr Val His Asp Leu
              50                55                60
Arg Leu Asn Glu Asn Lys Leu Lys Ala Val Leu Tyr Ser Ser Leu
              65                70                75
Asn Arg Phe Gly Asn Leu Thr Asp Leu Asn Leu Thr Lys Asn Glu
              80                85                90
Ile Ser Tyr Ile Glu Asp Gly Ala Phe Leu Gly Gln Ser Ser Leu
              95                100               105
Gln Val Leu Gln Leu Gly Tyr Asn Lys Leu Ser Asn Leu Thr Glu
              110               115               120
Gly Met Leu Arg Gly Met Ser Arg Leu Gln Phe Leu Phe Val Gln
              125               130               135
His Asn Leu Ile Glu Val Val Thr Pro Thr Ala Phe Ser Glu Cys
              140               145               150
Pro Ser Leu Ile Ser Ile Asp Leu Ser Ser Asn Arg Leu Ser Arg
              155               160               165
Leu Asp Gly Ala Thr Phe Ala Ser Leu Ala Ser Leu Met Val Cys
              170               175               180
Glu Leu Ala Gly Asn Pro Phe Asn Cys Glu Cys Asp Leu Phe Gly
```

27/46

```
                    185                 190                 195
Phe Leu Ala Trp Leu Val Val Phe Asn Asn Val Thr Lys Asn Tyr
                200                 205                 210
Asp Arg Leu Gln Cys Glu Ser Pro Arg Glu Phe Ala Gly Tyr Pro
                215                 220                 225
Leu Leu Val Pro Arg Pro Tyr His Ser Leu Asn Ala Ile Thr Val
                230                 235                 240
Leu Gln Ala Lys Cys Arg Asn Gly Ser Leu Pro Ala Arg Pro Val
                245                 250                 255
Ser His Pro Thr Pro Tyr Ser Thr Asp Ala Gln Arg Glu Pro Asp
                260                 265                 270
Glu Asn Ser Gly Phe Asn Pro Asp Glu Ile Leu Ser Val Glu Pro
                275                 280                 285
Pro Ala Ser Ser Thr Thr Asp Ala Ser Ala Gly Pro Ala Ile Lys
                290                 295                 300
Leu His His Val Thr Phe Thr Ser Ala Thr Leu Val Val Ile Ile
                305                 310                 315
Pro His Pro Tyr Ser Lys Met Tyr Ile Leu Val Gln Tyr Asn Asn
                320                 325                 330
Ser Tyr Phe Ser Asp Val Met Thr Leu Lys Asn Lys Lys Glu Ile
                335                 340                 345
Val Thr Leu Asp Lys Leu Arg Ala His Thr Glu Tyr Thr Phe Cys
                350                 355                 360
Val Thr Ser Leu Arg Asn Ser Arg Arg Phe Asn His Thr Cys Leu
                365                 370                 375
Thr Phe Thr Thr Arg Asp Pro Val Pro Gly Asp Leu Ala Pro Ser
                380                 385                 390
Thr Ser Thr Thr Thr His Tyr Ile Met Thr Ile Leu Gly Cys Leu
                395                 400                 405
Phe Gly Met Val Ile Val Leu Gly Ala Val Tyr Tyr Cys Leu Arg
                410                 415                 420
Lys Arg Arg Met Gln Glu Glu Lys Gln Lys Ser Val Asn Val Lys
                425                 430                 435
Lys Thr Ile Leu Glu Met Arg Tyr Gly Ala Asp Val Asp Ala Gly
                440                 445                 450
Ser Ile Val His Ala Ala Gln Lys Leu Gly Glu Pro Pro Val Leu
                455                 460                 465
Pro Val Ser Arg Met Ala Ser Ile Pro Ser Met Ile Gly Glu Lys
                470                 475                 480
Leu Pro Thr Ala Lys Gly Leu Glu Ala Gly Leu Asp Thr Pro Lys
                485                 490                 495
Val Ala Thr Lys Gly Asn Tyr Ile Glu Val Arg Thr Gly Ala Gly
                500                 505                 510
Gly Asp Gly Leu Ala Arg Pro Glu Asp Asp Leu Pro Asp Leu Glu
                515                 520                 525
Asn Gly Gln Gly Ser Ala Ala Glu Ile Ser Thr Ile Ala Lys Glu
                530                 535                 540
Val Asp Lys Val Asn Gln Ile Ile Asn Asn Cys Ile Asp Ala Leu
                545                 550                 555
Lys Leu Asp Ser Ala Ser Phe Leu Gly Gly Gly Ser Ser Ser Gly
                560                 565                 570
Asp Pro Glu Leu Ala Phe Glu Cys Gln Ser Leu Pro Ala Ala Ala
                575                 580                 585
Ala Ala Ser Ser Ala Thr Gly Pro Gly Ala Leu Glu Arg Pro Ser
                590                 595                 600
Phe Leu Ser Pro Pro Tyr Lys Glu Ser Ser His His Pro Leu Gln
```

```
                    605                 610                 615
Arg Gln Leu Ser Ala Asp Ala Ala Val Thr Arg Lys Thr Cys Ser
                    620                 625                 630
Val Ser Ser Ser Gly Ser Ile Lys Ser Ala Lys Val Phe Ser Leu
                    635                 640                 645
Asp Val Pro Asp His Pro Ala Ala Thr Gly Leu Ala Lys Gly Asp
                    650                 655                 660
Ser Lys Tyr Ile Glu Lys Gly Ser Pro Leu Asn Ser Pro Leu Asp
                    665                 670                 675
Arg Leu Pro Leu Val Pro Ala Gly Ser Gly Gly Gly Ser Gly Gly
                    680                 685                 690
Gly Gly Gly Ile His His Leu Glu Val Lys Pro Ala Tyr His Cys
                    695                 700                 705
Ser Glu His Arg His Ser Phe Pro Ala Leu Tyr Tyr Glu Glu Gly
                    710                 715                 720
Ala Asp Ser Leu Ser Gln Arg Val Ser Phe Leu Lys Pro Leu Thr
                    725                 730                 735
Arg Ser Lys Arg Asp Ser Thr Tyr Ser Gln Leu Ser Pro Arg His
                    740                 745                 750
Tyr Tyr Ser Gly Tyr Ser Ser Ser Pro Glu Tyr Ser Ser Glu Ser
                    755                 760                 765
Thr His Lys Ile Trp Glu Arg Phe Arg Pro Tyr Lys Lys His His
                    770                 775                 780
Arg Glu Glu Val Tyr Met Ala Ala Gly His Ala Leu Arg Lys Lys
                    785                 790                 795
Val Gln Phe Ala Lys Asp Glu Asp Leu His Asp Ile Leu Asp Tyr
                    800                 805                 810
Trp Lys Gly Val Ser Ala Gln Gln Lys Leu
                    815                 820
```

```
<210> 18
<211> 2653
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6431478CB1

<400> 18
gctgcggctg agcccagcgc tcgaggcgcg aggcagccag gagggcccgt gcggcgcggg 60
gagccagcga gcgcgccttc ggcattggcc gccgcgatgt cagctcagtg ctgtgcgggc 120
cagctggcct gctgctgtgg gtctgcaggc tgctctctct gctgtgattg ctgccccagg 180
attcggcagt ccctcagcac ccgcttcatg tacgccctct acttcattct ggtcgtcgtc 240
ctctgctgca tcatgatgtc aacaaccgtg gctcacaaga tgaaagagca cattcctttt 300
tttgaagata tgtgtaaagg cattaaagct ggtgacacct gtgagaagct ggtgggatat 360
tctgccgtgt atagagtctg tttttggaatg gcttgtttct tctttatctt ctgtctactg 420
accttgaaaa tcaacaacag caaaagttgt agagctcata ttcacaatgg cttttggttc 480
tttaaacttc tgctgttggg ggccatgtgc tcaggagctt tcttcattcc agatcaggac 540
acctttctga acgcctggcg ctatgtggga gccgtcggag gcttcctctt cattggcatc 600
cagctcctcc tgctcgtgga gtttgcacat aagtggaaca agaactggac agcaggcaca 660
gccagtaaca agctgtggta cgcctccctg gccctggtga cgctcatcat gtattccatt 720
gccactggag gcttggtttt gatggcagtg ttttatacac agaaagacag ctgcatggaa 780
aacaaaattc tgctgggagt aaatggaggc ctgtgcctgc ttatatcatt ggtagccatc 840
tcaccctggg tccaaaatcg acagccacac tcgggctctc tacaatcagg ggtcataagc 900
tgctatgtca cctacctcac cttctcagct ctgtccagca aacctgcaga agtagttcta 960
```

```
gatgaacatg ggaaaaatgt tacaatctgt gtgcctgact ttggtcaaga cctgtacaga 1020
gatgaaaact tggtgactat actggggacc agcctcttaa tcggatgtat cttgtattca 1080
tgtttgacat caacaacaag atcgagttct gacgctctgc aggggcgata cgcagctcct 1140
gaattggaga tagctcgctg ttgtttttgc ttcagtcctg gtggagagga cactgaagag 1200
cagcagccgg ggaaggaggg accacgggtc atttatgacg agaagaaagg caccgtctac 1260
atctactcct acttccactt cgtgttcttc ctagcttccc tgtatgtgat gatgaccgtc 1320
accaactggt tcaactacga aagtgccaac atcgagagct tcttcagcgg gagctggtcc 1380
atcttctggg tcaagatggc ctcctgctgg atatgcgtgc tgttgtacct gtgtacgctg 1440
gtcgctcccc tctgctgccc cacccgggag ttctctgtgt gatgatatcg gcggtcccct 1500
gggctttgtg ggcctacagc ctggaaagtg ccatcttttg aacagtgtcc ccggggcagg 1560
gactggcgcc ctgtgcctga gtgggtctga aaaagctttg agagagaaaa aaaaaaatct 1620
cctgattagc tttttacttt tgaaattcaa aaagaaacta ccagtttgtc ccaaaggaat 1680
tgaaattttc aaccaaactg atcatggttg aaatatctta cccctaggaa ctggatacca 1740
gttatgttga cttccttctg catgttttg ccaaaacaga atttggggca cagcatcttt 1800
tcacagggat aaaaatatct cgtgggccca gtcattctca tcctcggaat agaaaaacat 1860
gccaaaatct tgagtcccca cgcctaaca gaatccagac ccctctcact cacttccgcc 1920
tcttagagcc ttgtccccag ggggctttga ggacaggact cagcctgcag ggccctggt 1980
atttataggg tccaagagga ggcacctgct tttcaactgc accctcagtg ctgcctcttc 2040
acggccccta aacgtttccc tttgaggttg tgatgctggg aatcacagac ttcactctct 2100
gcctgcaccc ttccccgagg tctcatcttt tctgggtccc acatctttgt aataatgtga 2160
aaaagcacaa tttgtctgat cacccccag gtggttcccc accttattat cactacctga 2220
tccgagttac tgcaataagt acggcgctta tttatggtgt tagtcacatg attatagaac 2280
aagattcatg ttttctctgc ctaagcaatg gagggctatc attcttactt gtttgtgctg 2340
ttgataatga taatactttt aggaccttaa ctgaaaagct gcttcgtgtt gaagcctgct 2400
gcatgcactg ctctttcagt tgttgaggtc agccctcag ttttttctcc accttgaggc 2460
ctttgaaact gtaaaagcgg aagtcgtttt gtgttctgga tctgtaacgt gaccataccg 2520
ttcaggttca tgctggcatc cttggagtag atttgctaat gtgagaattt ctgaggtgag 2580
gatctcagac acactgacca gaagaagctt gttaggcaat gtgtggaagt ggccgaatat 2640
acttaaaaag agg                                                     2653
```

<210> 19
<211> 3531
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3584654CB1

<400> 19
```
ggaagcggtc ggggctgcac actcggatcg gcggggccgg ctcccgggcc cggccggctg 60
gaggagggag ggaaggaggc gggagggagc gagcggagcc atgggtgcgc acgtacgccc 120
cagcgctggg atttatcggc tcgcgaggag agcggagcag gcgcgcggcc caggcggagg 180
agcgccgact ctggagcagc cggagctgga agaggaggag gaggagaggc ggcggggaag 240
gaggaggagg gggagagtcg ctcccgccgg gcgagcatgg ggcgcctggc ctcgaggccg 300
ctgctgctgg cgctcctgtc gttggctctt tgccgagggc gtgtggtgag agtccccaca 360
gcgaccctgg ttcgagtggt gggcactgag ctggtcatcc cctgcaacgt cagtgactat 420
gatggcccca gcgagcaaaa ctttgactgg agcttctcat ctttggggag cagctttgtg 480
gagcttgcaa gcacctggga ggtggggttc ccagcccagc tgtaccagga gcggctgcag 540
aggggcgaga tcctgttaag gcggactgcc aacgacgccg tggagctcca cataaagaac 600
gtccagcctt cagaccaagg ccactacaaa tgttcaaccc ccagcacaga tgccactgtc 660
cagggaaact atgaggacac agtgcaggtt aaagtgctgg ccgactccct gcacgtgggc 720
cccagcgcgc ggccccgcc gagcctgagc ctgcgggagg gggagccctt cgagctgcgc 780
tgcaccgccg cctccgcctc gccgctgcac acgcacctgg cgctgctgtg ggaggtgcac 840
cgcggcccgg ccaggcggag cgtcctcgcc ctgacccacg agggcaggtt ccaccggggc 900
ctggggtacg agcagcgcta ccacagtggg gacgtgcgcc tcgacaccgt gggcagcgac 960
```

```
gcctaccgcc tctcagtgtc ccgggctctg tctgccgacc agggctccta caggtgtatc 1020
gtcagcgagt ggatcgccga gcagggcaac tggcaggaaa tccaagaaaa ggccgtggaa 1080
gttgccaccg tggtgatcca gccgacagtt ctgcgagcag ctgtgcccaa gaatgtgtct 1140
gtggctgaag gaaaggaact ggacctgacc tgtaacatca caacagaccg agccgatgac 1200
gtccggcccg aggtgacgtg gtccttcagc aggatgcctg acagcaccct acctggctcc 1260
cgcgtgttgg cgcggcttga ccgtgattcc ctggtgcaca gctcgcctca tgttgctttg 1320
agtcatgtgg atgcacgctc ctaccattta ctggttcggg atgttagcaa agaaaactct 1380
ggctactatt actgccacgt gtccctgtgg gcaccggac acaacaggag ctggcacaaa 1440
gtggcagagg ccgtgtcttc cccagctggt gtgggtgtga cctggctaga accagactac 1500
caggtgtacc tgaatgcttc caaggtcccc gggtttgcgg atgacccac agagctggca 1560
tgccgggtgg tggacacgaa gagtgggggag gcgaatgtcc gattcacggt ttcgtggtac 1620
tacaggatga accggcgcag cgacaatgtg gtgaccagcg agctgcttgc agtcatggac 1680
ggggactgga cgctaaaata tggagagagg agcaagcagc gggcccagga tggagacttt 1740
attttttcta aggaacatac agacacgttc aatttccgga tccaaaggac tacagaggaa 1800
gacagaggca attattactg tgttgtgtct gcctggacca aacagcggaa caacagctgg 1860
gtgaaaagca aggatgtctt ctccaagcct gttaacatat tttgggcatt agaagattcc 1920
gtgcttgtgg tgaaggcgag gcagccaaag ccttctttg ctgccggaaa tacatttgag 1980
atgacttgca aagtatcttc caagaatatt aagtcgccac gctactctgt tctcatcatg 2040
gctgagaagc ctgtcggcga cctctccagt cccaatgaaa cgaagtacat catctctctg 2100
gaccaggatt ctgtggtgaa gctggagaat tggacagatg catcacgggt ggatggcgtt 2160
gttttagaaa aagtgcagga ggatgagttc cgctatcgaa tgtaccagac tcaggtctca 2220
gacgcagggc tgtaccgctg catggtgaca gcctggtctc ctgtcagggg cagcctttgg 2280
cgagaagcag caaccagtct ctccaatcct attgagatag acttccaaac ctcaggtcct 2340
atatttaatg cttctgtgca ttcagacaca ccatcagtaa ttcggggaga tctgatcaaa 2400
ttgttctgta tcatcactgt cgagggagca gcactggatc cagatgacat ggcctttgat 2460
gtgtcctggt ttgcggtgca ctctttggc ctggacaagg ctcctgtgct cctgtcttcc 2520
ctggatcgga agggcatcgt gaccacctcc cggagggact ggaagagcga cctcagcctg 2580
gagcgcgtga gtgtgctgga attcttgctg caagtgcatg gctccgagga ccaggacttt 2640
ggcaactact actgttccgt gactccatgg gtgaagtcac caacaggttc ctggcagaag 2700
gaggcagaga tccactccaa gcccgttttt ataactgtga agatggatgt gctgaacgcc 2760
ttcaagtatc ccttgctgat cggcgtcggt ctgtccacgg tcatcgggct cctgtcctgt 2820
ctcatcgggt actgcagctc ccactggtgt tgtaagaagg aggttcagga gacacggcgc 2880
gagcgccgca ggctcatgtc gatggagatg gactaggctg gcccgggagg ggagtgacag 2940
agggacgttc taggagcaat tggggcaaga agaggacagt gatatttaa aacaaagtgt 3000
gttacactaa aaaccagtcc tctctaatct caggtgggac ttggcgctct ctcttttctg 3060
catgtcaagt tctgagcgcg gacatgttta ccagcacacg gctcttcttc ccacggcact 3120
ttctgatgta acaatcgagt gtgtgttttc ccaactgcag cttttaatg gttaaccttc 3180
atctaatttt ttttctccca ctggtttata gatcctctga cttgtgtgtg tttatagctt 3240
ttgtttcgcg gggttgtggt gaggaagggg tgatggcatg cggagttctt tgtcttcagt 3300
gagaatgtgc ctgccgcct gagagccagc ttccgcgttg gaggcacgtg ttcagagagc 3360
tgctgagcgc caccctctac ccggctgaca gacaacacag acctgtgccg aaggctaatt 3420
tgtggctttt acgaccctac cccacccct gttttcaggg gtttagacta catttgaaat 3480
ccaaacttgg agtatataac ttcttattga gcccaactgc tttttttttt t         3531
```

```
<210> 20
<211> 2280
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3737084CB1

<400> 20
gcgcgcccat ttcgagccca agtttccagc tcgggtttcc aggctcagaa ttttccagga 60
gtaggttctt gggcagtggc tgtgggagct ggaatggcgc agctggaagg ttactatttc 120
```

```
tcggccgcct tgagctgtac cttttttagta tcctgcctcc tcttctccgc cttcagccgg  180
gcgttgcgag agccctacat ggacgagatc ttccacctgc ctcaggcgca gcgctactgt  240
gagggccatt tctcccttttc ccagtgggat cccatgatta ctacattacc tggcttgtac  300
ctggtgtcaa ttggagtgat caaacctgcc atttggatct ttggatggtc tgaacatgtt  360
gtctgctcca ttgggatgct cagatttgtt aatcttctct tcagtgttgg caacttctat  420
ttactatatt tgcttttctg caaggtacaa cccagaaaca aggctgcctc aagtatccag  480
agagtcttgt caacattaac actagcagta tttccaacac tttattttttt taacttcctt  540
tattatacag aagcaggatc tatgtttttt actcttttttg cgtatttgat gtgtctttat  600
ggaaatcata aaacttcagc cttccttgga ttttgtggct tcatgtttcg gcaaacaaat  660
atcatctggg ctgtcttctg tgcaggaaat gtcattgcac aaaagttaac ggaggcttgg  720
aaaactgagc tacaaaagaa ggaagacaga cttccaccta ttaaaggacc atttgcagaa  780
ttcagaaaaa ttcttcagtt tcttttggct tattccatgt cctttaaaaa cttgagtatg  840
cttttgcttc tgacttggcc ctacatcctt ctgggatttc tgttttgtgc ttttgtagta  900
gttaatggtg gaattgttat tggcgatcgg agtagtcatg aagcctgtct tcattttcct  960
caactattct acttttttttc atttactctc ttttttttcct ttcctcatct cctgtctcct  1020
agcaaaatta agacttttct ttccttagtt tggaaacgta gaattctgtt ttttgtggtt  1080
accttagtct ctgtgttttt agtttggaaa ttcacttatg ctcataaata cttgctagca  1140
gacaatagac attatacttt ctatgtgtgg aaaagagttt ttcaaagata tgaaactgta  1200
aaatatttgt tagttccagc ctatatattt gctggttgga gtatagctga ctcattgaaa  1260
tcaaagtcaa ttttttttggaa tttaatgttt ttcatatgct tgttcactgt tatagttcct  1320
cagaaactgc tggaatttcg ttacttcatt ttaccttatg tcatttatag gcttaacata  1380
cctctgcctc ccacatccag actcatttgt gaactgagct gctatgcagt tgttaatttc  1440
ataacttttt tcatctttct gaacaagact tttcagtggc caaatagtca ggacattcaa  1500
aggtttatgt ggtaatatca gtgatatttc gaactgtgaa aatggactta ataattagac  1560
catttctaca aagaacaact gaataggtgg aaaacatgga atttcttttta ggtgcagtgg  1620
tggtcttcaa attacattag tttttttttat atatattttta aacatatgta agaaattaag  1680
tggcaaagaa ctgagaaagc ttaagacctg cttcaaaagc ctgaaaaatg gaaaaataaa  1740
attgtttttca gatatctcat atcactctca taatgttggc cccttaaaaa gcttgggaat  1800
gtttttgtatg tacaagttta ttaaaactgg gtatgcttca aaaaaaaaaa aaaaggggggg  1860
gggttcccac ccccaattcc gaaacctgga aaagcggttc cccgggggaaa attttttacc  1920
ccccccaaatt cccccaaaaa ttggggggcccg ggagcctaaa ggtactaccc cggggggggccc  1980
taaggggtgg gccccccccc attaattggg gtggcccaa tgccccgttt ccaattggga  2040
aacctttttgg tcccacccct tttattaatt ggccaacccc cggggaaaag gggtttttcct  2100
tttgggggcc ttttcccgttc cccggccaat aaaccggttc ccccgggtttt tcgggttcgg  2160
ggaagggttt ccagcccccc aaagggggggt aaacgggttt cccccaaaatt cggggggaaaa  2220
ccccggaaaa aacatttttg cccaaagggc ccccaaaagg ccaggccccct taaaaaggcc  2280
```

```
<210> 21
<211> 1104
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71426238CB1

<400> 21
taaagagagt tttgccttct tttgagccta agtcatgagt tggatgttcc tcagagatct   60
cctgagtgga gtaaataaat actccactgg gattggatgg atttggctgg ctgtcgtgtt  120
tgtcttccgt ttgctggtct acatggtggc agcagagcac gtgtggaaag atgagcagaa  180
agagtttgag tgcaacagta gacagcccgg ttgcaaaaat gtgtgttttg atgacttctt  240
ccccatttcc caagtcagac tttgggcctt acaactgata atggtctcca caccttcact  300
tctggtggtt ttacatgtag cctatcatga gggtagagag aaaaggcaca gaaagaaact  360
ctatgtcagc ccaggtacaa tggatggggg cctatggtac gcttatctta tcagcctcat  420
tgttaaaact ggttttgaaa ttggcttcct tgtttttattt tataagctat atgatggctt  480
tagtgttccc taccttataa agtgtgattt gaagccttgt cccaacactg tggactgctt  540
```

```
catctccaaa cccactgaga agacgatctt catcctcttc ttggtcatca cctcatgctt 600
gtgtattgtg ttgaatttca ttgaactgag tttttggtt ctcaagtgcc ttattaagtg 660
ctgtctccaa aaatatttaa aaaaacctca agtcctcagt gtgtgagtgc cacagcctca 720
gatatgttga atgtggtagg agagggaccc ctcccctact ccagaatctt cacacttggc 780
cataaacaca ctccctctac ctgaagcaaa gctactctgt gacacacaag agggttaaac 840
aaagaaaacc tgcatccctc ctcagcaagg cctaagctga gttggaagac aaagcacatc 900
agctttagta tcatttggga ggaatttttt tacattgtca atatgctttc agttatgagc 960
tctagacaga ggtctcattg ttttgttgta gggttctcca gtatgtggat aacattagtt 1020
gttttagaat aggtaattgc aaattagtct gaagaaatct aacaggattc ttttaagagc 1080
ttagattttt cagggaaaaa aaaa                                         1104
```

<210> 22
<211> 4966
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475123CB1

<400> 22
```
ggcggccgag ggcgattgcg gggcgcgcag gccgcgtgca cccgggacgc ttcccctcgg 60
ggaccctccg cgggcttctc cgccgcgccg tccggcggga gccggcggga ccccgggcga 120
gcggcgcggg cggcaccatg aggcggcagt ggggcgcgct gctgcttggc gccctgctct 180
gcgcacacgg cctggccagc agccccgagt gtgcttgtgg tcggagccac ttcacatgtg 240
cagtgagtgc tcttggagag tgtacctgca tccctgccca gtggcagtgt gatggagaca 300
atgactgcgg ggaccacagc gatgaggatg gatgtatact acctacctgt tcccctcttg 360
actttcactg tgacaatggc aagtgcatcc gcgctcctg ggtgtgtgac ggggacaacg 420
actgtgagga tgactcggat gagcaggact gtccccccg ggagtgtgag gaggacgagt 480
ttccctgcca gaatggctac tgcatccgga gtctgtggca ctgcgatggt gacaatgact 540
gtggcgacaa cagcgatgag cagtgtgaca tgcgcaagtg ctccgacaag gagttccgct 600
gtagtgacgg aagctgcatt gctgagcatt ggtactgcga cggtgacacc gactgcaaag 660
atggctccga tgaggagaac tgtccctcag cagtgccagc gccccctgc aacctggagg 720
agttccagtg tgcctatgga cgctgcatcc tcgacatcta ccactgcgat ggcgacgatg 780
actgtggaga ctggtcagac gagtctgact gctgtgagta ctctggccag ctgggagcct 840
cccaccagcc ctgccgctct ggggagttca tgtgtgacag tggcctgtgc atcaatgcag 900
gctggcgctg cgatggtgac gcggactgtg atgaccagtc tgatgagcgc aactgcaact 960
ggcagaccaa gtcaatccag cgtgttgaca aatactcagg ccggaacaag gagacagtgc 1020
tggcaaatgt ggaaggactc atggatatca tcgtggtttc ccctcagcgg cagacaggga 1080
ccaatgcctg tggtgtgaac aatggtggct gcacccacct ctgctttgcc agagcctcgg 1140
acttcgtatg tgcctgtcct gacgaacctg atagccggcc ctgctccctt gtgcctggcc 1200
tggtaccacc agctcctagg gctactggca tgagtgaaaa gagcccagtg ctacccaaca 1260
caccacctac caccttgtat tcttcaacca cccggacccg cacgtctctg gaggaggtgg 1320
aaggaaggat ggacatccgt cgaatcagct ttgacacaga ggacctgtct gatgatgtca 1380
tcccactggc tgacgtgcgc agtgctgtgg cccttgactg ggactcccgg gatgaccacg 1440
tgtactggac agatgtcagc actgatacca tcagcagggc caagtgggat ggaacaggac 1500
aggaggtggt agtggatacc agtttggaga gcccagctgg cctggccatt gattgggtca 1560
ccaacaaact gtactggaca gatgcaggta cagaccggat tgaagtagcc aacacagatg 1620
gcagcatgag aacagtactc atctgggaga accttgatcg tcctcgggac atcgtggtgg 1680
aacccatggg cgggtacatg tattggactg actggggtgc gagccccaag attgaacgag 1740
ctggcatgga tgcctcaggc cgccaagtca ttatctcttc taatctgacc tggcctaatg 1800
ggttagctat tgattatggg tcccagcgtc tatactgggc tgacgccggc atgaagacaa 1860
ttgaatttgc tggactggat ggcagtaaga ggaaggtgct gattggaagc cagctccccc 1920
acccatttgg gctgaccctc tatggagagc gcatctattg gactgactgg cagaccaaga 1980
gcatacagag cgctgaccgg ctgacagggc tggaccggga gactctgcag gagaacctgg 2040
aaaacctaat ggacatccat gtcttccacc gccgccggcc cccagtgtct acaccatgtg 2100
```

```
ctatggagaa tggcggctgt agccacctgt gtcttaggtc cccaaatcca agcggattca 2160
gctgtacctg ccccacaggc atcaacctgc tgtctgatgg caagacctgc tcaccaggca 2220
tgaacagttt cctcatcttc gccaggagga tagacattcg catggtctcc ctggacatcc 2280
cttattttgc tgatgtggtg gtaccaatca acattaccat gaagaacacc attgccattg 2340
gagtagaccc ccaggaagga aaggtgtact ggtctgacag cacactgcac aggatcagtc 2400
gtgccaatct ggatggctca cagcatgagg acatcatcac cacagggcta cagaccacag 2460
atgggctcgc ggttgatgcc attggccgga aagtatactg gacagacacg ggaacaaacc 2520
ggattgaagt gggcaacctg gacgggtcca tgcggaaagt gttggtgtgg cagaaccttg 2580
acagtccccg ggccatcgta ctgtaccatg agatggggtt tatgtactgg acagactggg 2640
gggagaatgc caagttagag cggtccggaa tggatggctc agaccgcgcg gtgctcatca 2700
acaacaacct aggatggccc aatggactga ctgtggacaa ggccagctcc caactgctat 2760
gggccgatgc ccacaccgag cgaattgagg ctgctgacct gaatggtgcc aatcggcata 2820
cattggtgtc accggtgcag cacccatatg gcctcaccct gctcgactcc tatatctact 2880
ggactgactg gcagactcgg agcatccacc gtgctgacaa gggtactggc agcaatgtca 2940
tcctcgtgag gtccaacctg ccaggcctca tggacatgca ggctgtggac cgggcacagc 3000
cactaggttt taacaagtgc ggctcgagaa atggcggctg ctcccacctc tgcttgcctc 3060
ggccttctgg cttctcctgt gcctgcccca ctggcatcca gctgaaggga gatgggaaga 3120
cctgtgatcc ctctcctgag acctacctgc tcttctccag ccgtggctcc atccggcgta 3180
tctcactgga caccagtgac cacaccgatg tgcatgtccc tgttcctgag ctcaacaatg 3240
tcatctccct ggactatgac agcgtggatg gaaaggtcta ttacacagat gtgttcctgg 3300
atgttatcag gcgagcagac ctgaacggca gcaacatgga gacagtgatc gggcgagggc 3360
tgaagaccac tgacgggctg gcagtggact gggtggccag gaacctgtac tggacagaca 3420
caggtcgaaa taccattgag gcgtccaggc tggatggttc ctgccgcaaa gtactgatca 3480
acaatagcct ggatgagccc cgggccattg ctgtttttcc caggaagggg tacctcttct 3540
ggacagactg gggccacatt gccaagatcg aacgggcaaa cttggatggt tctgagcgga 3600
aggtcctcat caacacagac ctgggttggc ccaatggcct taccctggac tatgataccc 3660
gcaggatcta ctgggtggat gcgcatctgg accggatcga gagtgctgac ctcaatggga 3720
aactgcggca ggtcttggtc agccatgtgt cccacccctt tgccctcaca cagcaagaca 3780
ggtggatcta ctggacagac tggcagacca agtcaatcca gcgtgttgac aaatactcag 3840
gccggaacaa ggagacagtg ctggcaaatg tggaaggact catggatatc atcgtggttt 3900
cccctcagcg gcagacaggg accaatgcct gtggtgtgaa caatggtggc tgcacccacc 3960
tctgctttgc cagagcctcg gacttcgtat gtgcctgtcc tgacgaacct gatagccagc 4020
cctgctccct tgtgcctggc ctggtaccac cagctcctag ggctactggc atgagtgaaa 4080
agagcccagt gctacccaac acaccaccta ccaccttgta ttcttcaacc acccggaccc 4140
gcacgtctct ggaggaggtg gaaggaagat gctctgaaag ggatgccagg ctgggcctct 4200
gtgcacgttc caatgacgct gttcctgctg ctccagggga aggacttcat atcagctacg 4260
ccattggtgg actcctcagt attctgctga ttttggtggt gattgcagct ttgatgctgt 4320
acagacacaa aaaatccaag ttcactgatc ctggaatggg gaacctcacc tacagcaacc 4380
cctcctaccg aacatccaca caggaagtga agattgaagc aatccccaaa ccagccatgt 4440
acaaccagct gtgctataag aaagagggag ggcctgacca taactacacc aaggagaaga 4500
tcaagatcgt agagggaatc tgcctcctgt ctggggatga tgctgagtgg gatgacctca 4560
agcaactgcg aagctcacgg ggggggcctcc tccgggatca tgtatgcatg aagacagaca 4620
cggtgtccat ccaggccagc tctggctccc tggatgacac agagacggag cagctgttac 4680
aggaagagca gtctgagtgt agcagcgtcc atactgcagc cactccagaa agacgaggct 4740
ctctgccaga cacgggctgg aaacatgaac gcaagctctc ctcagagagc caggtctaaa 4800
tgcccacatt ctcttccctg cctgcctgtt ccttctcctt tatggacgtc tagtccttgt 4860
gctcgcttac accgcaggcc ccgcttctgt gtgcttgtcc tcctcctcct cccacccat 4920
aactgttcct aagccttcac cggagctgtt taccacgtga gtcata            4966
```

<210> 23
<211> 5401
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7481952CB1

<400> 23

```
atggaccaga gcatcagcat tacctgggaa cttagtggaa atgcagaacc tcaggccctg 60
gcccagcctt acagaaccaa aagctacatg gaacaagcaa agcatctcac ctgtgacttt 120
gagtcgggtt tctgcggttg ggagccattt ctcacagaag attcacactg gaagctgatg 180
aaaggattga ataatggaga gcaccacttt cctgcagctg atcacacagc aaacataaat 240
catggatcgt ttatttattt ggaggcacag cgctcccccg gggtggccaa gcttggaagt 300
cctgttctta caaaattgct cactgcctct accccatgtc aggtgcagtt ttggtatcat 360
ttgtctcaac attcaaatct ctcagttttt acaagaacgt ctctagatgg aaacttgcaa 420
aagcagggca aaataatcag attctccgaa tctcagtgga gccacgcaaa aattgatctc 480
attgcagaag cgggagaatc tactctacct tttcagttaa ttttggaagc tactgttttg 540
tcgtcaaatg ctaccgttgc tctagatgac atcagtgtgt cccaggaatg tgaaatttcc 600
tataaatcac taccaaggac cagtacacaa agcaagtttt ccaagtgtga ctttgaagca 660
aacagctgtg attggtttga agtaattagt ggtgaccatt ttgactggat acggagctct 720
cagagtgaac tttctgctga ttttgagcac caggctccac ctcgggatca tagtctcaac 780
gcatctcaag ggcattttat gttcattctg aagaaaagca gcagcttgtg gcaagttgct 840
aagcttcaga gcccaacttt cagccagaca ggacctggat gcatactttc cttctggttc 900
tataactatg gcctgtcagt gggagcagct gagctgcagc tacatatgga aaattctcat 960
gactcaacag tgatttggag agtattatac aatcagggca aacaatggtt ggaggcaacc 1020
attcagctag ggcgcctttc gcagcccttc catttgtcac tagataaagt cagtctgggc 1080
atttatgatg gggtctcagc tattgatgac atccgatttg aaaattgtac tctccctctt 1140
cctgctgaga gctgtgaagg gctggatcat ttctggtgtc gccacaccag ggcttgcata 1200
gaaaagcttc ggttatgtga tctggtggat gactgtggtg atcgtactga tgaagtcaac 1260
tgtgcacctg agctgcagtg taactttgaa actggaatct gtaactggga acaagatgca 1320
aaagatgact ttgattggac caggaaccag ggtccaactc caacacttaa cacagggcca 1380
atgaaagata cactctggg cacagctaaa ggacactatc tctacataga atcttcagag 1440
ccacaggctt ttcaagacag tgctgcctta ctcagcccaa tccttaatgc cactgataca 1500
aaaggctgca ccttccgctt ctattaccac atgtttggaa agcgcatta taggttggca 1560
atctaccaac gaatctggag tgactcaagg ggacagctgc tgtggcagat atttgggaat 1620
caaggcaaca gatggattag gaaacacctc aacatttcca gcaggcagcc ctttcagata 1680
ttggtggagg cttcagtggg agatggcttc actggagata ttgcgattga tgatctgtca 1740
tttatggact gcaccctcta ccctggtaat ttgccagcag acctcccaac tccaccagaa 1800
acgtcagttc ctgtaacatt acctccacac aactgcacag acagtgaatt tatctgcagg 1860
tctgatggtc actgcattga aaaaatgcag aaatgtgatt ttaaatatga ctgccctgac 1920
aaatcagatg aagcatcctg tgttatggaa gtttgcagct ttgagaaaag aagcctgtgt 1980
aaatggtatc aaccaatccc agtacatttg cttcaagatt caaacacatt caggtggggg 2040
cttgggaacg ggatcagcat tcatcatggg gaagaaaacc acaggccatc agtggatcat 2100
acacaaaata ccactgatgg ctggtacctg tatgctgaca gttctaatgg gaaatttggt 2160
gacacggctg acattctcac tcctatcatt tcactcacgg gaccaaaatg taccttggtg 2220
ttctggacac atatgaatgg ggccaccgtt ggttctctcc aggtgctcat caagaaagat 2280
aacgttactt ctaaattgtg ggctcaaact ggacagcaag gtgcacagtg aagagagca 2340
gaagtgtttt taggcattcg ttcacataca cagattgtct tcagagccaa acgtggtatc 2400
agttacatag gagatgtagc agtggatgat atttccttcc aagattgctc ccctttgctt 2460
agcccagaga gaaagtgtac tgatcatgaa ttcatgtgtg ctaataagca ctgcattgcc 2520
aaagacaagc tgtgtgattt tgtgaatgat tgtgctgata attcagatga gactactttc 2580
atttgccgta cctccagtgg gcgctgtgat ttcgaatttg atctttgttc ctggaagcag 2640
gagaaagatg aggactttga ctggaacctg aaagctagca gcatccctgc agcaggcaca 2700
gagccagcag cagatcacac tttgggaaat tcatctggtc attacatctt tataaagagt 2760
ttgtttcctc agcagcccat gagagctgcc agaatttcaa gtccagttat aagtaagaga 2820
agcaaaaact gcaagattat ttttcattat cacatgtatg gaaatggcat tggggcactc 2880
accttaatgc aggtgtcagt cacaaaccaa acgaaggttc tacttaacct cactgtagaa 2940
caaggcaatt tctggcggag agaagaactg tcactgtttg gtgatgaaga cttccaactc 3000
aaatttgaag gtagagttgg gaaaggtcag cgtggagaca ttgcacttga tgacattgtg 3060
cttacagaaa attgtctatc actccatgat tccgtgcaag aagaactggc agtgcctctt 3120
ccaacaggtt tctgcccact tggctatagg gaatgtcata atggaaaatg ctataggctg 3180
```

```
gaacaaagct gtaacttcgt agataactgt ggagataata ctgatgaaaa tgagtgtggt 3240
agctcctgta cttttgaaaa aggctggtgt ggctggcaaa actcccaggc tgacaacttt 3300
gattgggttt tagggghttgg ctctcatcaa agcttaagac ctcccaaaga ccacacactt 3360
ggaaatgaaa atgggcactt catgtatctg gaagctactg cagtgggcct tcggggtgac 3420
aaagcacact tcaggagtac catgtggcga gaatccagtg cagcctgcac catgagcttc 3480
tggtatttca tatctgcaaa ggccacagga tccattcaga ttctcatcaa gacagagaaa 3540
ggactatcaa aagtatggca agaaagtaag cagaaccctg gtaatcattg gcaaaaggct 3600
gacatcctgc taggaaagtt aaggaatttt gaagtcatat ttcaaggtat cagaacaagg 3660
gacctgggag gaggagctgc aattgatgat attgaattta aaaactgcac aactgtggga 3720
gagatctctg agctttgtcc ggaaatcact gatttttgt gccgggacaa gaagtgcatt 3780
gcatcccacc ttctttgtga ctataagcca gactgctctg ataggtctga tgaagctcac 3840
tgtgcacatt atacaagcac aacaggaagc tgcaattttg aaacaagttc aggaaactgg 3900
accacagcct gcagtcttac tcaagactct gaggatgact tggactgggc cattggcagc 3960
agaattcctg ccaaagcatt aattccagac tctgatcaca cgccaggtag tggtcagcac 4020
ttcctgtacg tcaactcatc tggctccaag gaaggatccg ttgccagaat tactacttcc 4080
aaatccttcc cagcaagcct tggaatgtgt actgttcggt tctggttcta catgattgat 4140
cccaggagta tgggaatatt aaaggtgtat accattgaag aatcggggct aaacatcctg 4200
gtgtggtcag tgattggaaa taaaagaacg ggatggacat atggctctgt gcctctctcc 4260
agtaacagtc cgtttaaggt ggcatttgaa gctgatttgg atggaaatga ggacatcttt 4320
attgctcttg atgacatctc ttttacccca gagtgtgtga ctggaggtcc tgtcccagtg 4380
cagccatcac cctgtgaagc tgatcagttt tcttgtatct acacactcca atgtgtccct 4440
ctctcaggga aatgtgatgg acatgaagac tgcatagatg gatctgatga aatggattgt 4500
cctctcagcc ccaccctcc actctgtagt aacatggagt tcccgtgctc tacagacgag 4560
tgtatacctt ccctcctgct atgcgatgga gtgccgact gccactttaa tgaagatgag 4620
ctcatctgct ccaacaaaag ctgttctaat ggagctctgg tgtgtgcctc ctccaacagc 4680
tgtatcccag cccaccagcg ctgtgatggt tttgccgact gcatggattt ccagcttgat 4740
gagtccagct gctcagaatg tccattaaat tactgcagaa atggtgggac ttgtgtagtg 4800
gagaaaaatg gtcctatgtg tcgatgtaga caaggctgga aggaaatcg atgccatatc 4860
aagtttaatc ctcctgctac agacttcaca tacgctcaga ataatacatg gactctcctg 4920
ggtattggat tagcattcct gatgactcac atcacagttg cagtcttgtg tttcttgca 4980
aacagaaagg taccaataag gaaaaccgag ggaagtggta actgtgcctt tgtcaatcca 5040
gtttacggga ctggagcaa cccagagaaa acagagagtt ctgtctattc cttctcaaac 5100
ccattatatg gcacaacatc aggaagcctg gagaccctgt cacatcatct caaatagcag 5160
catcgagacc aagtctgatc caacatgtgt agtttctaga aaattgaagt ctccacaatc 5220
tgatagaaac tcatcttcta caatggtaaa aagagaaagg attgtaaatg ccagtgtaat 5280
tataacattt atgaatgaat tttcttgcag aatatagaga atgtttatat ggaatcagaa 5340
tcagtacctt atcttcactg aacatctgaa tattttaata aaatttctat ttaatcaaaa 5400
a                                                               5401
```

```
<210> 24
<211> 1949
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 382654CB1

<400> 24
aggcagaggg ctaggtggaa aaagcattga aggccatgag atggctgtga gagagaacaa 60
aggggcagaa gtgcacagag ctactgtggg ggaggagata gcacccaggc ttaagaagcc 120
aggattggca gggagtgaag agccagagag gcgaagcttt gggagatcag agggcttaaa 180
gttgggagtg ggctaaggag cccaggcct gatgcttccc tcttcctcat gggcctctgt 240
tcacagaccc cctggagggg gtgaacatca ccagccccgt gcgcctgatc catggcaccg 300
tggggaagtc ggctctgctt tctgtgcagt acagcagtac cagcagcgac aggcctgtag 360
tgaagtggca gctgaagcgg gacaagccag tgaccgtggt gcagtccatt ggcacagagg 420
```

```
tcatcggcac cctgcggcct gactatcgag accgtatccg actctttgaa aatggctccc 480
tgcttctcag cgacctgcag ctggccgatg agggcaccta tgaggtcgag atctccatca 540
ccgacgacac cttcactggg gagaagacca tcaaccttac tgtagatgtg cccatttcga 600
ggccacaggt gttggtggct tcaaccactg tgctggagct cagcgaggcc ttcaccttga 660
actgctcaca tgagaatggc accaagccca gctacacctg gctgaaggat ggcaagcccc 720
tcctcaatga ctcgagaatg ctcctgtccc ccgaccaaaa ggtgctcacc atcacccgcg 780
tgctcatgga ggatgacgac ctgtacagct gcatggtgga gaaccccatc agccagggcc 840
gcagcctgcc tgtcaagatc accgtataca gaagaagctc cctttacatc atcttgtcta 900
caggaggcat cttcctcctt gtgaccttgg tgacagtctg tgcctgctgg aaaccctcca 960
aaaggaaaca gaagaagcta gaaaagcaaa actccctgga atacatggat cagaatgatg 1020
accgcctgaa accagaagca gacaccctcc ctcgaagtgg tgagcaggaa cggaagaacc 1080
ccatggcact ctatatcctg aaggacaagg actccccgga gaccgaggag aacccggccc 1140
cggagcctcg aagcgcgacg gagcccggcc cgcccggcta ctccgtgtct cccgccgtgc 1200
ccggccgctc gccggggctg cccatccgct ctgccgccg ctacccgcgc tccccagcgc 1260
gctccccagc caccggccgg acacactcgt cgccgcccag ggccccgagc tcgcccggcc 1320
gctcgcgcag cgcctcgcgc acactgcgga ctgcgggcgt gcacataatc cgcgagcaag 1380
acgaggccgg cccggtggag atcagcgcct gagccgcctc gggatcccct gagaggcgcc 1440
cgcggtctgc ggccagtggc ccggggggaaa gctggggctg ggaagcccgg gcgcggcgcg 1500
ctggggacga ggggaggtcc cggggggggcg ctggtgtctc gggtgtgaac gtgtatgagc 1560
atgcgcagac ggaggcgggt gcgcggaggc ggcagtgttg atatggtgaa accgggtcgc 1620
atttgcttcc ggtttactgg ctgtgtcctc acttggtata ggttgtgcca tggggttctt 1680
ccgttcctgc tcaccacttc gagggagggt gtctgcttct ggtttcaggc ggtcatcatt 1740
ctgatccatg tattccaggg agtttcgctt ttctagcttc ttctgtttcc ttttggaggg 1800
tttccagcag gcacagactg tcaccaaggt cacaaggagg aagatgcctc ctgtagacaa 1860
gatgatgtac agggagcttc tcctagggag agagagaagc gagagcagga gggcctcccg 1920
gggccagatg tgtgaccact gccctacta                                    1949
```

<210> 25
<211> 2133
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1867351CB1

<220>
<221> unsure
<222> 2110
<223> a, t, c, g, or other

<400> 25
```
cactgccggc ctccgcggta cccactgccg gcctccgcgc tacccggccg cagcgcgcga 60
gtcacatgga agctcctgag gagcccgcgc cagtgcgcgg aggcccggag gccacccttg 120
aggtccgtgg gtcgcgctgc ttgcggctgt ccgccttccg agaagagctg cgggcgctct 180
tggtcctggc tggccccgcg ttcttggttc agctgatggt gttcctgatc agcttcataa 240
gctccgtgtt ctgtggccac ctgggcaagc tggagctgga tgcagtcacg ctggcaatcg 300
cggttatcaa tgtcactggt gtctcagtgg gattcggctt atcttctgcc tgtgacaccc 360
tcatctccca gacgtacggg agccagaacc tgaagcacgt gggcgtgatc ctgcagcgga 420
gtgcgctcgt cctgctcctc tgctgcttcc cctgctgggc gctctttctc aacacccagc 480
acatcctgct gctcttcagg caggacccag atgtgtccag gcttacccag acctatgtca 540
cgatcttcat tccagctctt cctgcaacct ttctttatat gttacaagtt aaatatttgc 600
tcaaccaggg aattgtactg ccccagatcg taactggagt tgcagccaac cttgtcaatg 660
ccctcgccaa ctatctgttt ctccatcaac tgcatcttgg ggtgataggc tctgcactgg 720
caaacttgat ttcccagtac accctggctc tactcctctt tctctacatc ctcgggaaaa 780
aactgcatca agctacatgg ggaggctggt ccctcgagtg cctgcaggac tgggcctcct 840
```

```
tcctccgcct ggccatcccc agcatgctca tgctgtgcat ggagtggtgg gcctatgagg 900
tcgggagctt ccccagtggc atcctcggca tggtggagct gggcgctcag tccatcgtgt 960
atgaactggc catcattgtg tacatggtcc ctgcagactt cagtgtggct gccagtgtcc 1020
gggtaggaaa cgctctgggt gctggagaca tggagcaggc acggaagtcc tctaccgttt 1080
ccctgctgat tacagtgctc tttgctgtag ccttcagtgt cctgctgtta agctgtaagg 1140
atcacgtggg gtacattttt actaccgacc gagacatcat taatctggtg gctcaggtgg 1200
ttccaattta tgctgtttcc cacctctttg aagctcttgc ttgcacgagt ggtggtgttc 1260
tgaggggggag tggaaatcag aaggttggag ccattgtgaa taccattggg tactatgtgg 1320
ttggcctccc catcgggatc gcgctgatgt ttgcaaccac acttggagtg atgggtctgt 1380
ggtcagggat catcatctgt acagtctttc aagctgtgtg ttttctaggc tttattattc 1440
agctaaattg gaaaaaagcc tgtcagcagg ctcaggtaca cgccaatttg aaagtaaaca 1500
acgtgcctcg gagtgggaat tctgctctcc ctcaggatcc gcttcaccca gggtgccctg 1560
aaaaccttga aggaatttta acgaacgatg ttggaaagac aggcgagcct cagtcagatc 1620
agcagatgcg ccaagaagaa cctttgccgg aacatccaca ggacggcgct aaattgtcca 1680
ggaaacagct ggtgctgcgg cgagggcttc tgctcctggg ggtcttctta atcttgctgg 1740
tggggatttt agtgagattc tatgtcagaa ttcagtgacg tggtaggaaa gaaagtcagg 1800
tcaagtgatg ctttttgagct tacacacaat tcacaggccc accagtgaca atttactgtg 1860
agttaatgtc attcaggtgt gcccatggat tttgagggct ggaaatgcaa agacacattt 1920
ttctataaaa agaaaaagca actaaggtta aaagctatat tgtggcccaa gacactgttt 1980
gtgaaagatg .ccatgattag taattcacca ctatcttgaa ccaagcacag gatcaatgtg 2040
ctgactgcat cggccaatgg ctttgatact tctgctattt ttttagacac aacccataac 2100
tacggggatn actagttcta agcgccggca ccg                             2133
```

```
<210> 26
<211> 2090
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> misc_feature
<223> Incyte ID No: 3323104CB1
```

```
<400> 26
ggcggaagcg agaccgtcca tccagaggaa ggcaagtttt tggctcgggc ggctgagaag 60
accgcgcggg gctggagaca ggtagcagta cggggcgggg gcttcatgcc ggatgtgata 120
gtctgcagtc gtttcggttg gcagcctggc gggtgggaga tgcggcggcc acctgctgca 180
aagaaccgaa gggaaggtta gaagtacgaa ggcagtttgg agctggggct aagcagctgt 240
cgcacggtca gatcatgggc tccaccaagc actggggcga attgctcctg aacttgaagg 300
tggctccagc cggcgtcttt ggtgtggcct ttctagccag agtcgccctg gtttttctatg 360
gcgtcttcca ggaccggacc ctgcacgtga ggtatacgga catcgactac caggtcttca 420
ccgacgccgc gcgcttcgtc acggaggggc gctcgcctta cctgagagcc acgtaccgtt 480
acacccccgct gctgggttgg ctcctcactc ccaacatcta cctcagcgag ctctttggaa 540
agtttctctt catcagctgc gacctcctca ccgctttcct cttataccgc ctgctgctgc 600
tgaaggggct ggggcgccgc caggcttgtg gctactgtgt ctttttggctt cttaacccc 660
tgcctatggc agtatccagc cgcggtaatg cggactctat tgtcgcctcc ctggtcctga 720
tggtcctcta cttgataaag aaaagactcg tcgcgtgtgc agctgtattc tatggtttcg 780
cggtgcatat gaagatatat ccagtgactt acatccttcc cataaccctc cacctgcttc 840
cagatcgcga caatgacaaa agcctccgtc aattccggta cactttccag gcttgtttgt 900
acgagctcct gaaaaggctg tgtaatcggg ctgtgctgct gtttgtagca gttgctggac 960
tcacgttttt tgccctgagc tttggtttttt actatgagta cggctgggaa tttttttggaac 1020
acacctactt ttatcacctg actaggcggg atatccgtca caacttttct ccgtacttct 1080
acatgctgta tttgactgca gagagcaagt ggagtttttc cctgggaatt gctgcattcc 1140
tgccacagct catcttgctt tcagctgtgt ctttcgccta ttacagagac ctcgtttttt 1200
gttgttttct tcatacgtcc attttttgtga cttttaacaa agtctgcacc tcccagtact 1260
ttctttggta cctctgctta ctgcctcttg tgatgccact agtcagaatg ccttggaaaa 1320
gagctgtagt tctcctaatg ttatggttaa tagggcaggc catgtggctg gctcctgcct 1380
```

```
atgttctaga gtttcaagga aagaacacct ttctgtttat ttggttagct ggtttgttct 1440
ttcttcttat caattgttcc atcctgattc aaattatttc ccattacaaa gaagaacccc 1500
tgacagagag aatcaaatat gactagtgta tgttccacac cctctgctac tgtgttacat 1560
tctgattgtc ttgtatggac cagaagagag ctttgggaca tttttttctga acattctaag 1620
cattctagtg aaagttccca tgttccaaca gaacttaaaa gcaatgtttg ccttatatat 1680
aaaagggaca caataattga ggtccacctt ctaggaaatc ctaggactcg tttatttggg 1740
acatggtggg aataaaggtc acatattgga aaatggaaag gctgatgaaa ctatcagata 1800
ctaaaacatt cttaaaatag aggaatatag ttagagacat caggtttaag ccagtatttg 1860
ttcctgtttt acaatgcttc tgtcttaagc tgtgtcttaa ctttttaacac ccatcttttc 1920
tttctaaagc tttcctgaca gctgtgaaaa tccaaaaaat attcttaaac tgtgtatggg 1980
ggcccttgcc tgtagtctca gcactttggg aggctgaggt gggagggtcg cttgagttca 2040
ggagttctag acccacctgg ggcaagatgg tgagacctag tctcaaaaaa        2090
```

<210> 27
<211> 1618
<212> DNA
<213> Homo sapiens

<220>   ·
<221> misc_feature
<223> Incyte ID No: 4769306CB1

<400> 27
```
agaatctatg ggattttcag ctcgatacaa tttcacacct gatcctgact ttaaggacct 60
tggagctttg aaaccattac cagcgtgtga gtttgagatg ggcggttccg aaggaattgt 120
ggagtctata caaattatga aggaaggcaa agctactgct agcgaggctg ttgattgcaa 180
gtggtacatc cgagcacctc cacggtccaa gatttactta cgattcttgg actatgagat 240
gcagaattca aatgagtgca agaggaattt tgtggctgtg tatgatggaa gcagttccgt 300
ggaggatttg aaagctaagt tctgtagcac tgtggctaat gatgtcatgc tacgcacggg 360
tcttggggtg atccgcatgt gggcagatga gggcagtcga aacagccgat ttcagatgct 420
cttcacatcc tttcaagaac ctccttgtga aggcaacaca ttcttctgcc atagtaacat 480
gtgtattaat aatactttgg tctgcaatgg actccagaac tgtgtgtatc cttgggatga 540
aaatcactgt aaagagaaga ggaaaaccag cctgctggac cagctgacca acaccagtgg 600
gactgtcatt ggcgtgactt cctgcatcgt gatcatcctc attatcatct ctgtcatcgt 660
acagatcaaa cagcctcgta aaaagtatgt ccaaaggaaa tcagactttg accagacagt 720
tttccaggag gtatttgaac ctcctcatta tgagttatgc actctcagag ggacaggagc 780
tacagctgac tttgcagatg tggcagatga ctttgaaaat taccataaac tgcggaggtc 840
atcttccaaa tgcattcatg accatcactg tggatcacag ctgtccagca ctaaaggcag 900
ccgcagtaac ctcagcacaa gagatgcttc tatcttgaca gagatgccca cacagccagg 960
aaaacccctc atcccaccca tgaacagaag aaatatcctt gtcatgaaac acaactactc 1020
gcaagatgct gcagatgcct gtgacataga tgaaatcgaa gaggtgccga ccaccagtca 1080
caggctgtcc agacacgata aagccgtcca gcggttctgc ctcattgggt ctctaagcaa 1140
acatgaatct gaatacaaca caactagggt ctagaaagaa aattcaagac agcttgagaa 1200
tagtgcgttc ctgaatgatt ttgaacatgc tacagtgaaa agtgacagtg tggaccatgg 1260
aatcaccagc tagagatgag gaaactgaag agttttagta acttttttaa gattacacaa 1320
taaacaatga tgaatcaagc tttgaagcca acctcaccaa ccacaagatc aaccaacact 1380
cttcaccaat gtgtaatata accacgttaa tattcaacat agtacgtact gctgaaagaa 1440
gttgatactt attcatatta accccgtagt tttgtgtttc ctcatctgta aaagtatgta 1500
ttataacacc ttctctccac cttacagcgt gtgaggttca aatgaccatt cattggaaga 1560
tattttttat atcctataat gcattataaa aataaatcat ttttcctaaa aaaaaaaa    1618
```

<210> 28
<211> 3269
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2720058CB1

<400> 28

```
cctgagggga atggctttcg tcccctcct cttggtgacc tggtcgtcag ccgccttcat 60
tatctcctac gtggtcgccg tgctctccgg gcacgtcaac cccttcctcc cgtatatcag 120
tgatacggga acaacacctc cagagagtgg tattttggga tttatgataa acttctctgc 180
atttcttggt gcagccacga tgtatacaag atacaaaata gtacagaagc aaaatcaaac 240
ctgctatttc agcactcctg tttttaactt ggtgtcttta gtgcttggat tggtgggatg 300
tttcggaatg ggcattgtcg ccaattttca ggagttagct gtgccagtgg ttcatgacgg 360
gggcgctctt ttggcctttg tctgtggtgt cgtgtacacg ctcctacagt ccatcatctc 420
ttacaaatca tgtccccagt ggaacagtct ctcgacatgc cacatacgga tggtcatctc 480
tgccgtttct tgcgcagctg tcatccccat gattgtctgt gcttcactaa tttccataac 540
caagctggag tggaatccaa gagaaaagga ttatgtatat cacgtagtga gtgcgatctg 600
tgaatggaca gtggcctttg gttttatttt ctacttccta actttcatcc aagatttcca 660
gagtgtcacc ctaaggatat ccacagaaat caatggtgat atttgaagaa agaagaattc 720
agtctcactc agtgaatgtc gcaggccatt tctaaaagtg ctacagagga cagacagggt 780
tttgaggcca ccctgattat tgggatgcat ctgcagcaca tccaggactt gaatttcatt 840
acgagttcct aatagttgta tttctaaaga tgtgtttcct agagaatgta cagccttatg 900
acactgtagt gatgttttta taattttcta agtagatttt tttatattaa caaattcata 960
tacagaaaaa ataaggtgtt acaaaaaatg gagagctctt attttttgtac agattctgtc 1020
gtttttgttt tatttgtgtg agatttatgg aaatacacta aatgagtaat tcaggttcag 1080
tacatttatt acaaagtgaa atcagggggat attcatttgt aaattttatt cttagtgaat 1140
gaactgtata attttttta tcaggagagc acttataaaa ttcaatttat aaagatcata 1200
tacccaaatc ataaagattt agttgataca ttaacactaa gatactctga tttttagccg 1260
aactaaacaa agtgcttcta ctgagaggcc tttataccac catgtacagt aactctaagt 1320
gaatacggaa gaccttggtt ttgaaattct gccaccttgt ttctccctgc tcatgaggtc 1380
gcaccttttg ctcttgctgc taattgccca ttcgtagtgg gtgtaatgcc aggtggaatg 1440
gtttcaacaa gtcaggtgaa aaccatcctt tattgttgct ggcacaactt gatatatagt 1500
ctgactcaga actgaagctc acatctcaaa ttcatttcat gccagtaaat gtggcaaaga 1560
gaagaaaggc ccaagagcga gacaagaaga atggagaagg gggcagccaa gaagaacttc 1620
tgggttcagg gtactgttta tttgctcctt ctcttcatgc ctgtggctgg atgtcccaca 1680
acactataag aaataagtca agccctttgt gttaagcaag aactacagac tccatctttt 1740
cacccaaatc atgaatgacc aataaaaagc aagttattcc agaggaagaa gcagcccttg 1800
aaatgttaag gcttaggctt gaaaggtgaa gagcaggaat tctctctttc aaatcctaga 1860
gcataaaccc atgtgtggcc aagtgagatc agccctcaag ggcacatgcc aagggcagag 1920
cagcccatgt agacagcttc ggagggcatg ggggtgtagg gagttcgggg tagctcctca 1980
ttaactattt gttgggtgag taaagggggtg aggctcagtg gcaggtacct ctgcaatgac 2040
aagctgcctc ccctctatgt gtttagcata tgttattaga acatgtccga caccctacc 2100
gctgccattt gggcccttta ataaagccaa gtagagaaat ctggcaataa aaggcaaatg 2160
taagcatgct ttcttaaga cgcatcataa atggtttct ttaagtgaat ggaagagttt 2220
gacagagata cacctttgta agaaaacatt aagaatgctg gctggctgtg gtggctcaca 2280
cctgtattcc cagcactttg ggaggcctag gcaggaggat tgcttgagcc tgggacttcg 2340
agaccagact gggaaacatg gcaaaatccc atctctacaa caaaaataca aaaattagcc 2400
aagtgcggtg gtgtgcctgt agtcctagtt acttgggagg ctgaggtggg agaatcacct 2460
gagcccagga ggtggaggct gcagtgagcc atgccaatgc actccagtct gggcaacaga 2520
gtgagaccct gtctcaaaaa taaataaata aataaatgaa taaagagaat gctaatcatt 2580
tctgggttca ctgcgactca ctgtagtgct ggggatcccc cttgtaacac tggaactgaa 2640
agacagtgat gaaagctatg tcaagcattc attattctga agaggaggag aaatgccaca 2700
tacctttccc atgggacctg tggtggaatg aatccatact tctgcctcac ttcgagcaga 2760
cttttgttct cggcgctcct cacgatggag tttcatgctt cattttcaca tctctctgca 2820
caattagatt gggagctcct tgagggcaga gtacgtgcct taatctttat ctttgtaatg 2880
ccacaatgaa cagagtgcct cctggtacac tgtaggagct taagaaatac tcactgaatg 2940
catgaatgaa tgaatgaaca aatgaaggaa tgactaagga tgtttgtagt gctataatat 3000
agaatgggat ttactctgct ttaccagtta gtttcataat aaacaaatag tctgtaacag 3060
```

```
aacattctgt acctgccata caggctcatg ttcatgccaa ttcttcctag agccaaataa 3120
ataaagactt aggggggggcc cccgaaaaaa gggccgccgc cccgggggttt atctccggcc 3180
cgggcctgaa gccgacaccg gttccccaag ggtacagctt tccccttggg ggactcaggg 3240
gaacagggtt ccccgggggca atttttacc                                 3269
```

<210> 29
<211> 1227
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481255CB1

<400> 29
```
atggacaggg ccaagcagca gcaggcgctg ctcctcctcc ctgtctgcct cgccctcacc 60
ttctccctca ccgccgtggt cagcagccac tggtgtgagg ggacccgacg ggtggtgaag 120
ccactgtgcc aggaccagcc gggagggcag cactgcattc acttcaaacg ggacaacagc 180
agcaatggca ggatggacaa caatagccag gctgtcctgt acatttggga gctgggtgat 240
gacaagttca ttcagcgggg gttccatgtg gggctctggc agtcctgcga ggagagcctc 300
aacggtgaag atgaaaagt taggagtttc cggagtgtag tgccagctga agaacaaggt 360
gttttgtggc tgtccatcgg gggcgaggtc ctggatatcg ttctgatact gacaagcgcc 420
atcctcctgg gctccagagt gagttgtcgc agccctgggt tccactggct cagggtggat 480
gccttggtag ccatcttcat ggtgctggca gggcttctag gcatggtggc ccacatgatg 540
tacacaacca tttttcaaat cactgtgaac cttggaccag aagattggaa gcctcagacc 600
tgggactatg gctggtcata ttgccttgcc tggggttctt tcgccctctg cctggctgtg 660
tcggtctcgg ccatgagcag gttcacggca gcccgcctgg aattcaccga gaagcagcag 720
gcacagaacg gcagtcggca ctctcaacac agcttcctgg aacccgaggc ttcggagagc 780
atttggaaaa caggagctgc tccttgccct gctgaacaag ccttcaggaa tgtttctgga 840
cacctcccac caggcgcccc aggcaaggtg tccatatgct agccagtgtc catggctgcc 900
acatccgcac aggcaaacaa gccaggcact gacactcaca atgtacaccc tgcctctggg 960
ttggacttca ggagattgtt gtccagggaa agcttccatc cccaccctc cacatctcac 1020
cttttactaa acacctttgg ctttagcctt tgattcctgt taaaatgcca gtaccttgaa 1080
gtgagataat gcttactgaa gatatcaacc attgacactc tagtataaaa gagagcttct 1140
taatgacagt gaatttgata aggataccaa agaaacaggg aggatgccag tactaaggga 1200
agagaagttg aagaaagagg aaagcaa                                    1227
```

<210> 30
<211> 2618
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1510242CB1

<400> 30
```
agcctaatac cttctcaagt tgatctcccc ccaggcacag ccttgtccca gccggaagac 60
tcaaatttta aaatttcgaa ttctgaatag tttattcatg tatataagtt actgacacag 120
taagagggat ctttttttta tcttacaaga cctaaaaatt acttaatacc tttgaaataa 180
aatgttttat ttctgccaaa tggcattatg aatataataa gacttaagag caccaaaagt 240
tagttactac agcaagatac actagtatac gtatatctat ttatattaag aaactcaggg 300
cacttgtcta taattcacaa gttaccaatc ttaaacattt aaggcgaccg ccgcgagtcc 360
gcagtagttc gggccatgga ggcggagccg ccgctctacc cgatggcggg ggctgcgggg 420
ccgcagggcg acgaggacct gctcgggggtc ccggacgggc ccgaggcccc gctggacgag 480
ctggtgggcg cgtaccccaa ctacaacgag gaggaggagg agcgccgcta ctaccgccgc 540
```

```
aagcggcctg ggcgtgctca agaacgtgct ggctgccagc gccgggggca tgctcaccta 600
cggcgtctac ctgggcctcc tgcagatgca gctgatcctg cactacgacg agacctaccg 660
cgaggtgaag tatggcaaca tggggctgcc cgacatcgac agcaaaatgc tgatgggcat 720
caacgtgact cccatcgccg ccctgctcta cacacctgtg ctcatcaggt tttttggaac 780
gaagtggatg atgttcctcg ctgtgggcat ctacgccctc tttgtctcca ccaactactg 840
ggagcgctac tacacgcttg tgccctcggc tgtggccctg ggcatggcca tcgtgcctct 900
ttgggcttcc atgggcaact acatcaccag gatggcgcag aagtaccatg agtactccca 960
ctacaaggag caggatgggc aggggatgaa gcagcggcct ccgcggggct cccacgcgcc 1020
ctatctcctg gtcttccaag ccatcttcta cagcttcttc catctgagct tcgcctgcgc 1080
ccagctgccc atgatttatt tcctgaacca ctacctgtat gacctgaacc acacgctgta 1140
caatgtgcag agctgcggca ccaacagcca cgggatcctc agcggcttca acaagacggt 1200
tctgcggacg ctcccgcgga gcggaaacct cattgtggtg gagagcgtgc tcatggcagt 1260
ggccttcctg gccatgctgc tggtgctggg tttgtgcgga gccgcttacc ggcccacgga 1320
ggagatcgat ctgcgcagcg tgggctgggg caacatcttc cagctgccct tcaagcacgt 1380
gcgtgactac cgcctgcgcc acctcgtgcc tttctttatc tacagcggct tcgaggtgct 1440
ctttgcctgc actggtatcg ccttgggcta tggcgtgtgc tcggtggggc tggagcggct 1500
ggcttacctc ctcgtggctt acagcctggg cgcctcagcc gcctcactcc tgggcctgct 1560
gggcctgtgg ctgccacgcc cggtgcccct ggtggccgga gcagggggtgc acctgctgct 1620
caccttcatc ctcttttttct gggcccctgt gcctcgggtc ctgcaacaca gctggatcct 1680
ctatgtggca gctgcccttt ggggtgtggg cagtgccctg aacaagactg gactcagcac 1740
actcctggga atcttgtacg aagacaagga gagacaggac ttcatcttca ccatctacca 1800
ctggtggcag gctgtggcca tcttcaccgt gtacctgggc tcgagcctgc acatgaaggc 1860
taagctggcg gtgctgctgg tgacgctggt ggcggccgcg gtctcctacc tgcggatgga 1920
gcagaagctg cgccggggcg tggccccgcg ccagccccgc atcccgcggc cccagcacaa 1980
ggtgcgcggt taccgctact tggaggagga caactcggac gagagcgacg cggagggcga 2040
gcatggggac ggcgcggagg aggaggcgcc gcccgcaggg cccaggcctg ccccgagcc 2100
cgctggactc ggccgccggc cctgcccgta cgaacaggcg caggggggag acgggccgga 2160
ggagcagtga ggggccgcct ggtccccgga ctcagcctcc ctcctcgccg gcctcagttt 2220
accacgtctg aggtcggggg gacccctcc gagtcccgcg ctgtcttcaa aggccctgt 2280
ctccctcccc ccacgttggg gacgcccctc ccagagccca ggtcacctcc gggcttccgc 2340
agcccctcc aaggcggagt ggagccttgg gaaccctcg gccaagcaca ggggttcgaa 2400
aatacagctg aaaccccgcg ggcccttagc acgcgcccca cgccggagc acggtcaggg 2460
tcttcttgcg acccggcccg ctccagatcc ccacagctct cggccgcgga cccgggccgc 2520
gtgtgagcgc actttgcacc tcctatcccc agggtccgcc gagagccacg atttttttaca 2580
gaaaatgagc aataaagaga ttttgtactg tcaaaaaa                          2618
```

<210> 31
<211> 2188
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 162131CB1

<400> 31

```
ccggaattga ccaactggta gactcgccta gaggggacgc attgtgtcct agttgaggct 60
aacagtcagt atccagcctc aacattcagc agaggcccca gatcagcgtc tgagccaggc 120
caacaatgac caaggaggat gggatcctgg gtgcagctca tcacaagcgt cggggtgcag 180
caaaaccatc caggctggac agtggctgga cagttccaag aaaagaaacg cttcactgaa 240
gaagtcattg aatacttcca gaagaaagtt agcccagtgc atctgaaaat cctgctgact 300
agcgatgaag cctggaagag attcgtgcgt gtggctgaat gcccagggga agaagcagat 360
gctctctatg aagctctgaa gaatcttaca ccatatgtgg ctattgagga caagacatg 420
cagcaaaaag aacagcagtt tagggagtgg ttttgaaag agtttcctca aatcagatgg 480
aagattcagg agtccataga aaggcttcgt gtcattgcaa atgagattga aaaggtccac 540
agaggctgcg tcatcgccaa tgtggtgtct ggctccactg gcatcctgtc tgtcattggc 600
```

```
gttatgttgg caccatttac agcagggctg agcctgagca ttactgcagc tggggtaggg 660
ctgggaatag catctgccac ggctgggatc gcctccagca tcgtggagaa cacatacaca 720
aggtcagcag aactcacagc cagcaggctg actgcaacca gcactgacca attggaggca 780
ttaagggaca ttctgcatga catcacaccc aatgtgcttt cctttgcact tgattttgac 840
gaagccacaa aaatgattgc gaatgatgtc catacactca ggagatctaa agccactgtt 900
ggacgccctt tgattgcttg gcgatatgta cctataaatg ttgttgagac actgagaaca 960
cgtggggccc ccaccatggat agtgagaaaa gtagcccgga acctgggcaa ggccacttca 1020
ggtgtcctcg ttgtgctgga tgtagtcaac cttgtgcaag actcactgga cttgcacaag 1080
ggggaaaaat ccgagtctgc tgagttgctg aggcagtggg ctcaggagct ggaggagaat 1140
ctcaatgagc tcacccatat ccatcagagt ctaaaagcag gctaggccca attgttgcgg 1200
gaagtcaggg acccccaaacg gagggactgg ctgaagccat ggcagaagaa cgtggattgt 1260
gaagatttca tggacatta  ttagttcccc aaattaatac ttttataatt tcctatgcct 1320
gtctttaccg caatctctaa acacaaattg tgaagatttc atggacactt atcacttccc 1380
caatcaatac ccttgtgatt cttatgcct  gtctttactt taatctccta atcctgtcag 1440
ctgaggagga tgtatgtcac ctcaggacca tgtgataatt gcgttaactg cacaaattgt 1500
agagcatgtg tgtttgaaca atatgaaatc tgggcacctt gaaaaaagaa caggataaca 1560
gcaattgttc agggaataag agagataacc ttaaactctg accaacagtg agcctggtgg 1620
aacagagtca tatttctctt ctttcaaaag caaatgggag aaatatcgct gaattctttt 1680
tctcagcaag gaacatccct gagaaagaga atgcacccct gagggtgggt ctataaatgg 1740
cctccttggg tgtggccatc ttctatggtc gagactgtag ggatgaaata aaccccagtc 1800
tcccatagtg ctcccaggct tattaggaag aggaaattcc cgcctaataa attttggtca 1860
gaccggttgc tctcaaaacc ctgtctcctg ataagatgtt atcaatgaca atggtgcctg 1920
aaacctcatt agcaatttta atttctcccc ggtcctgtgg tcctgtgatc tcaccctgcc 1980
tccacttgcc ttgtgatatt ctattacctt gtgaagtagg tgatctttgt gacccacacc 2040
cacaccctat tcatacactc cctccccttt tgaaagtccc taataaaaac ttgctggttt 2100
tgcagcttgt gaggcatcac ggaacctacc gatgtgtgat gtctcccctg gacacctagc 2160
tttaaaattt ctaaaaaaaa aaaaaaaa                                    2188
```

<210> 32
<211> 1969
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1837725CB1

<400> 32
```
gtagcagcgg cggtccagtc gtagcccggc cgcccgcgcc tgtccggtcc ggtccggcca 60
cggaggcagc gcagcggcgg gactccgagc ctaccccgcc gagtgagctg cgccgcaccg 120
tgccgtccca cccggcaccc accagtccga tggggccgca gcggcggctg tcccctgccg 180
gggccgccct actctggggc ttcctgctcc agctgacagc cgctcaggaa gcaatcttgc 240
atgcgtctgg aaatggcaca accaaggact actgcatgct ttataaccct tattggacag 300
ctcttccaag taccctagaa aatgcaactt ccattagttt gatgaatctg acttccacac 360
cactatgcaa ccttttctgat attcctcctg ttggcataaa gagcaaagca gttgtggttc 420
catgggggaag ctgccatttt cttgaaaaag ccagaattgc acagaaagga ggtgctgaag 480
caatgttagt tgtcaataac agtgtcctat ttcctccctc aggtaacaga tctgaatttc 540
ctgatgtgaa aatactgatt gcatttataa gctacaaaga ctttagagat atgaaccaga 600
ctctaggaga taacattact gtgaaaatgt attctccatc gtggcctaac tttgattata 660
ctatggtggt tatttttgta attgcggtgt tcactgtggc attaggtgga tactggagtg 720
gactagttga attggaaaac ttgaaagcag tgacaactga agatagagaa atgaggaaaa 780
agaaggaaga atatttaact tttagtcctc ttacagttgt aatatttgtg gtcatctgct 840
gtgttatgat ggtcttactt tatttcttct acaaatggtt ggtttatgtt atgatagcaa 900
ttttctgcat agcatcagca atgagtctgt acaactgtct tgctgcacta attcataaga 960
taccatatgg acaatgcacg attgcatgtc gtggcaaaaa catggaagtg agacttattt 1020
ttctctctgg actgtgcata gcagtagctg ttgtttgggc tgtgtttcga aatgaagaca 1080
```

```
ggtgggcttg gattttacag gatatcttgg ggattgcttt ctgtctgaat ttaattaaaa 1140
cactgaagtt gcccaacttc aagtcatgtg tgatacttct aggccttctc ctcctctatg 1200
atgtattttt tgtttcata acaccattca tcacaaagaa tggtgagagt atcatggttg 1260
aactcgcagc tggacctttt ggaaataatg aaaagttgcc agtagtcatc agagtaccaa 1320
aactgatcta tttctcagta atgagtgtgt gcctcatgcc tgtttcaata ttgggtttttg 1380
gagacattat tgtaccaggc ctgttgattg catactgtag aagatttgat gttcagactg 1440
gttcttctta catatactat gtttcgtcta cagttgccta tgctattggc atgatactta 1500
catttgttgt tctggtgctg atgaaaaagg ggcaacctgc tctcctctat ttagtacctt 1560
gcacacttat tactgcctca gttgttgcct ggagacgtaa ggaaatgaaa aagttctgga 1620
aaggtaacag ctatcagatg atggaccatt tggattgtgc aacaaatgaa gaaaaccctg 1680
tgatatctgg tgaacagatt gtccagcaat aatattatgt ggaactgcta taatgtgtca 1740
ttgattttct acaaatagac ttcgactttt taaattgact tttgaattga caatctgaaa 1800
gagtcttcaa tgatatgctt gcaaaaatat attttatga gctggtactg acagttacat 1860
cataaataac taaaacgctt tgctttaat gttaaagttg tgccttcaca ttaaataaaa 1920
catatggtct gtgtagtttc cgaaaaaaaa aaaaaaaaa aaaaaaaaa        1969
```

```
<210> 33
<211> 3006
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3643847CB1

<400> 33
gccatgcagg cggcgcgcgt ggactacatc gctccctggt gggtcgtgtg gctgcacagc 60
gtcccgcacg tcggcctgcg cctgcagccc gtgaacagca ccttcagccc cggcgacgag 120
agttaccagg agtcgctgct gttcctgggg ctggtggccg ccgtctgcct gggcctgaac 180
ctcatcttcc ttgtggctta cctggtctgt gcatgccact gccggcggga cgatgcggtg 240
cagaccaagc agcaccactc ctgctgcatc acctggacgg ccgtggtggc cgggctcatc 300
tgctgtgctg cggtgggcgt tggtttctat ggaaacagcg agaccaacga tggggcgtac 360
cagctgatgt actccttgga cgatgccaac cacaccttct ctgggatcga tgctctggtt 420
tccggaacta cccagaagat gaaggtggac ctagagcagc acctggcccg gctcagtgag 480
atctttgctg cccgggggcga ttacctgcag accctgaagt tcatacagca gatggcgggc 540
agcgttgttg ttcagctctc aggactgccc gtgtggaggg aggtcaccat ggagctgacc 600
aagctatccg accagactgg ctacgtggag tactacaggt ggctctccta cctcctgctc 660
tttatcctgg acctggtcat ctgcctcatt gcctgcctgg gactggccaa gcgctccaag 720
tgtctcctgg cctcgatgct gtgctgtggg gcactgagcc tgctcctcag ttgggcatcc 780
ctggccgctg atggctctgc ggcagtggcc accagtgact tctgtgtggc tcctgacacc 840
ttcatcctga acgtcacgga gggccagatc agcacagagg tgactcgcta ctacctgtat 900
tgcagccaga gtggaagcag ccccttccag cagaccctga ccacttcca gcgcgcactc 960
accaccatgc agatccaggt cgcgggggctg ctgcagtttg ccgtgcccct cttctccact 1020
gcagaggaag acctgcttgc aatccagctc ctgctgaact cctcagagtc cagccttcac 1080
cagctgactg ccatggtgga ctgccgaggg ctgcacaagg attatctgga cgctcttgct 1140
ggcatctgct acgacggcct ccagggcttg ctgtaccttg gctcttctc cttcctggcc 1200
gccctcgcct tctccaccat gatctgtgcg gggccaaggg cctggaagca cttcaccacc 1260
agaaacagag aatacgatga cattgatgat gatgacccct ttaacccca agcctggcgc 1320
atggcggctc acagtccccc gaggggacag cttcacagct tctgcagcta cagcagtggc 1380
ctgggaagtc agaccagcct gcagcccccg gcccagacca tctccaacgc ccctgtctcc 1440
gagtacatga accaagccat gctctttggt aggaacccac gctacgagaa cgtgccacta 1500
atcgggagag cctcccctcc gcctacgtac tctcccagca tgagagccac ctacctgtct 1560
gtggcggatg agcacctgag gcactacggg aatcagtttc agcctaaca gactttcggg 1620
ggttcctgcc tcctttttcc gttctggttt ttaattagtg caaatacaag ctgcgtttct 1680
ttaatagaaa ccaaaggcat ctggagcccg agaggcctcc tgctgtggca gaggagcagc 1740
tgggattccc gaccaaagcc caggggggtg cagaagactc accacgcggg ccagcctctc 1800
```

```
tcttttgccc tgctctccac accagaaatg cccccaggtg cttggctgcc tcagaggtac 1860
catccctgag ctggctgcct ggccctgctc acccctacgc ctcgcccttg ccaggagggg 1920
agtggcagtg aggagggggc caggtcaggc accaccatca agagagctgt gtgttctctc 1980
tggtcccaca acgatgactc tgcctcttgt cagcccagcc aagagcccag acgacccctc 2040
tgtcctcgtt ccctgtcctc gttccctgca ggtaacatga gaagggctga tcaggagatg 2100
ctctttaaga agttcgcacc cctgctgaca ccagaacagc ccaaatcaga gttcccaggg 2160
ccagacaggc tcttcctggg ccacagaggg gaggcatcag gaaagctctg cagtgggggg 2220
ctggtggctc cggggctggg ggatcacagg ctggtgaacc ccggtgggaa cagaggtgaa 2280
agcctgccac attccgcctg tctccctaac cctccattgc ctcgcctcta ttccagaatc 2340
aatgctgcag aatgtgttag ctgcagatag gcatggtctc aggtatgaac agacactttg 2400
aaacgacttt aggtctttct tttctccagt gtttttaaaca tgttgattat ccaaagaatt 2460
gaaactccta gcacatccag tttttacaac agatttgcag ctcattcctt accctggtta 2520
ggtcactact tttgcagatt ttgctggcac tgatctggag atctgcagat ctggaggaga 2580
cgggaaggag tcgattctta aataaggatc agtgaggcat cctgtcccaa gctactgttt 2640
ggtggggatc tgggttcatc tcacccacag agggaggatc tttaagagga gaaaaaagcc 2700
aagagggaaa gccagagttc cctgttctag gggactagcc aaatgcctac atcagctgtc 2760
ccctccctgt tgtctccaag taagtttgcc agaaaaggtt ttagcaaagt gctacaactg 2820
tgtctttata ggaggatagg cctctgccct gccccacccc caccacctgt ccccaccag 2880
tgtcccaggc cacaggagct tattggccag gagggaataa tgtcccccaa tactgcctgt 2940
tgagggacca gagttggggt ctttggtgct tccaacctcc tgccaacctg gagttcacaa 3000
caccag                                                             3006
```

<210> 34
<211> 2884
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6889872CB1

<400> 34
```
tcactcctgg ctcagtgcgg cactctccag cctcctgtgg gaatcatctg aagttctgag 60
cccggaagcc aaggaggaag acgaggagga ggaggaggag gaggaggagg aggaggaggg 120
agaggaagtc aagccctgag aacccttgca ccttcctagc aggagacaag gagcaacgct 180
gcggtgggga gcaggctgtg gggcccccac ccccagccct agccaggcct agtgcctgct 240
gtagcaccct agaagatccc cagcagttgg cactagctgt acccaccttg cctggggccc 300
ccgtgctggg ggtcgccccc aagatggtgg cggccccagg gaggactgta ctgccagccc 360
cagcctctgg ccgctaggca ccccctgcct tgccctggcc cctcactccg aggccagcgc 420
catgctgcgc ctggggctgt gcgcggcggc gctgctgtgc gtgtgccggc cgggtgccgt 480
gcgtgccgac tgctggctca ttgagggcga caagggctac gtgtggctgg ccatctgcag 540
ccagaaccag ccgccctacg agaccatccc gcagcacatc aatagcaccg tgcacgacct 600
gcggctcaac gagaacaagc tcaaagccgt gctctactcc tcgctcaacc gctttgggaa 660
cctcaccgac ctcaacctca ccaagaacga gatctcctac atcgaggacg gtgccttcct 720
gggccagtcg agcctgcagg tcctgcagct gggctacaac aagctcagca acctgacgga 780
gggcatgctg cgaggcatga gccgcctgca gttcctcttt gtccagcaca acctcatcga 840
ggtggtgacg cccaccgcct tctccgagtg cccgagcctc atcagcatcg acctgtcctc 900
caaccgcctc agccgcctgg acggtgccac ctttgccagc ctcgccagcc tgatggtgtg 960
tgagctggcc ggcaacccct tcaactgtga gtgcgacctc ttcggcttcc tggcctggct 1020
ggtggtcttc aacaacgtca ccaagaacta cgaccgcctg cagtgtgagt cgccgcggga 1080
gtttgccggc tacccgctgc tggtgccccg gccctaccac agcctcaacg ccatcaccgt 1140
actccaggcc aagtgtcgga atggctcgct gccgccgcgg cccgtgagcc accccacgcc 1200
ctactccacc gacgcccaga gggagccaga cgagaactcg ggcttcaacc ccgacgagat 1260
cctttcggtg gagccgccgg cctcgtccac cacggatgcg tcggcagggc cagccatcaa 1320
gctgcaccac gtcacgttca cctcggccac cctggtggtc atcattccac accccacag 1380
caagatgtac atcctcgtgc agtacaacaa cagctacttc tccgacgtca tgaccctcaa 1440
```

```
gaacaagaag gagatcgtga cgctggacaa actgcgggcg cacactgagt acaccttctg 1500
cgtgacctcg ctgcgcaaca gccgccgctt caaccacacc tgcctgacct tcaccacgcg 1560
ggaccccgtc cccggagact tggcgcccag cacctccacc accacccact acatcatgac 1620
catcctgggc tgcctcttcg gcatggttat cgtgctggga gccgtgtact actgcctgcg 1680
caagcggcgc atgcaggagg agaagcagaa gtctgtcaac gtcaagaaga ccatcctgga 1740
gatgcgctac ggggctgatg tggatgccgg ctccattgtg cacgccgccc agaagctggg 1800
cgagcctccc gtgctgcccg tatctcgcat ggcctccatc ccctccatga tcggggagaa 1860
gctgcccacc gccaagggt tggaggccgg gctggacaca cccaaggtag ccaccaaagg 1920
caactatata gaggtgcgca caggcgccgg cggggacggt ctggctcggc ccgaggatga 1980
cctcccggac ctcgagaacg gccagggctc ggctgcagag atctccacca ttgccaagga 2040
ggtggacaag gtcaaccaga tcattaacaa ctgcatcgat gctctcaagc tggactcggc 2100
ctcttttctg ggaggcggca gcagcagtgg ggaccccgag ctggccttcg agtgccagtc 2160
cctccctgca gctgctgccg cctcctcagc cactggcccc gggggccctgg agcggcccag 2220
cttcctttcg cctccctaca aggagagctc ccaccaccca ctacagcgcc agctgagcgc 2280
cgacgcggcc gtgacccgca agacctgcag cgtgtcgtcc agtggttcca tcaagagcgc 2340
caaggtcttt agcctggacg tgcccgacca tccggccgcc acagggctgg ctaagggcga 2400
ctccaagtac atcgagaagg gcagccccct caacagcccg ctggaccggc tcccgctggt 2460
gccggcgggc agcggcgggg gcagcggcgg gggcgggggc atccaccacc tggaggtgaa 2520
gccggcctac cactgcagcg agcaccggca cagctttccc gccctgtact acgaggaggg 2580
tgccgacagc ctgagccagc gcgtgtcctt cctcaagccg ctgacccgct ccaagcgtga 2640.
ctccacctac tcgcagctct cccccagaca ctactactca gggtactcct ccagccccga 2700
gtactcatcc gagagcacgc acaagatctg ggagcgcttc cggccctaca agaagcacca 2760
ccgggaggag gtgtacatgg ccgccggtca cgccctgcgc aagaaggtcc agttcgccaa 2820
ggacgaggat ctgcatgaca tccttgatta ctggaagggg gtctccgccc agcagaagct 2880
gtga                                                                2884
```